



Separability between Signal and Noise Components Using the Distribution of Scaled Hankel Matrix Eigenvalues with Application in Biomedical Signals

Nader Alharbi

A Thesis Submitted in Partial Fulfillment of the Requirements of
Bournemouth University for the Degree of

Doctor of Philosophy

September 2017

Bournemouth University

Copyright

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with its author and due acknowledgement must always be made of the use of any material contained in, or derived from, this thesis.

Abstract

Biomedical signals are records from human and animal bodies. These records are considered as nonlinear time series, which hold important information about the physiological activities of organisms, and embrace many subjects of interest. However, biomedical signals are often corrupted by artifacts and noise, which require separation or signal extraction before any statistical evaluation. Another challenge in analysing biomedical signals is that their data is often non-stationary, particularly when there is an abnormal event observed within the signal, such as epileptic seizure, and can also present chaotic behaviour. The literature suggests that distinguishing chaos from noise continues to remain a highly contentious issue in the modern age, as it has been historically. This is because chaos and noise share common properties, which in turn make them indistinguishable. We seek to provide a viable solution to this problem by presenting a novel approach for the separability between signal and noise components and the differentiation of noise from chaos.

Several methods have been used for the analysis of and discrimination between different categories of biomedical signals, but many of these are based on restrictive assumptions of the normality, stationarity and linearity of the observed data. Therefore, an improved technique which is robust in its analysis of non-stationary time series is of paramount importance in accurate diagnosis of human diseases. The SSA (Singular Spectrum Analysis) technique does not depend on these assumptions, which could be very helpful for analysing and modelling biomedical data. Therefore, the main aim of the thesis is to provide a novel approach for developing the SSA technique, and then apply it to the analysis of biomedical signals.

SSA is a reliable technique for separating an arbitrary signal from a noisy time series (signal+noise). It is based upon two main selections: window length, L ; and the number of eigenvalues, r . These values play an important role in the reconstruction and forecasting stages. However, the main issue in extracting signals using the SSA procedure lies in identifying the optimal values of L and r required for signal reconstruction. The aim of this thesis is to develop theoretical and methodological aspects of the SSA technique, to present a novel approach to distinguishing between

deterministic and stochastic processes, and to present an algorithm for identifying the eigenvalues corresponding to the noise component, and thereby choosing the optimal value of r relating to the desired signal for separability between signal and noise. The algorithm used is considered as an enhanced version of the SSA method, which decomposes a noisy signal into the sum of a signal and noise. Although the main focus of this thesis is on the selection of the optimal value of r , we also provide some results and recommendations to the choice of L for separability. Several criteria are introduced which characterise this separability.

The proposed approach is based on the distribution of the eigenvalues of a scaled Hankel matrix, and on dynamical systems, embedding theorem, matrix algebra and statistical theory. The research demonstrates that the proposed approach can be considered as an alternative and promising technique for choosing the optimal values of r and L in SSA, especially for biomedical signals and genetic time series.

For the theoretical development of the approach, we present new theoretical results on the eigenvalues of a scaled Hankel matrix, provide some properties of the eigenvalues, and show the effect of the window length and the rank of the Hankel matrix on the eigenvalues. The new theoretical results are examined using simulated and real time series. Furthermore, the effect of window length on the distribution of the largest and smallest eigenvalues of the scaled Hankel matrix is also considered for the white noise process. The results indicate that the distribution of the largest eigenvalue for the white noise process has a positive skewed distribution for different series lengths and different values of window length, whereas the distribution of the smallest eigenvalue has a different pattern with L ; the distribution changes from left to right when L increases. These results, together with other results obtained by the different criteria introduced and used in this research, are very promising for the identification of the signal subspace.

For the practical aspect and empirical results, various biomedical signals and genetics time series are used. First, to achieve the objectives of the thesis, a comprehensive study has been made on the distribution, pattern; and behaviour of scaled

Hankel matrix eigenvalues using simulated data, including symmetric and nonsymmetric distributions, trend and sine wave series. The study includes comparison between the distribution of eigenvalues, the effect of the length of the series on the mean, the variance, and the coefficient of variation of the eigenvalues. Furthermore, the normal distribution with different parameters is considered and the effect of scale and shape parameters are evaluated. The correlation between eigenvalues is also assessed, using parametric and non-parametric association criteria. In addition, the distribution of eigenvalues for synthetic time series generated from some well known low dimensional chaotic systems are analysed in-depth. The results yield several important properties with broad application, enabling the distinction between chaos and noise in time series analysis. At this stage, the main result of the simulation study is that the findings related to the series generated from normal distribution with mean zero (white noise process) are totally different from those obtained for other series considered in this research, which makes a novel contribution to the area of signal processing and noise reduction.

Second, the proposed approach and its criteria are applied to a number of simulated and real data with different levels of noise and structures. Our results are compared with those obtained by common and well known criteria in order to evaluate, enhance and confirm the accuracy of the approach and its criteria. The results indicate that the proposed approach has the potential to split the eigenvalues into two groups; the first corresponding to the signal and the second to the noise component. In addition, based on the results, the optimal value of L that one needs for the reconstruction of a noise free signal from a noisy series should be the median of the series length. The results confirm that the performance of the proposed approach can improve the quality of the reconstruction step for signal extraction.

Finally, the thesis seeks to explore the applicability of the proposed approach for discriminating between normal and epileptic seizure electroencephalography (EEG) signals, and filtering the signal segments to make them free from noise. Various criteria based on the largest eigenvalue are also presented and used as features to distinguish between normal and epileptic EEG segments. These features can be considered as useful information to classify brain signals. In addition, the approach

is applied to the removal of nonspecific noise from *Drosophila* segmentation genes. Our findings indicate that when extracting signal from different genes, for optimised signal and noise separation, a different number of eigenvalues need to be chosen for each gene.

Contents

Copyright	i
Abstract	ii
Publications	x
List of Figures	xi
List of Tables	xiv
Abbreviations	xvi
Symbols	xvii
Acknowledgements	xix
Declaration of Authorship	xx
1 Introduction	1
1.1 Overview of singular spectrum analysis and separability	1
1.2 Biomedical signals and chaos	7
1.3 Aims, objectives, contributions and organisation of the thesis	12
2 Literature Review of SSA	16
2.1 Literature review	16
2.2 A Review of Applications in Biomedical Signals and Genetic Studies with SSA	22
2.2.1 Applications in biomedical signals with SSA	22
2.2.2 Applications in genetic studies with SSA	29
2.3 Summary	33
3 Theoretical Framework	34
3.1 White noise	34

3.2	Singular spectrum analysis	35
3.2.1	Embedding	35
3.2.2	Singular value decomposition	36
3.2.3	Grouping	38
3.2.4	Diagonal averaging	40
3.3	Separability	43
3.4	Determinants of a matrix	48
3.5	Theoretical results	49
3.5.1	The problem	49
3.5.2	The solution	54
3.5.3	The general form of ζ_i	60
3.5.4	The eigenvalues and determinant of \mathbf{B} for $L = 2$	63
3.5.5	The effect of the window length and rank of the Hankel matrix	65
3.5.6	Examples for evaluation of the theoretical results	68
3.5.6.1	The effect of L on the eigenvalues	68
3.5.6.2	Asymptotical behaviour of $I_{\mathbf{A}}$ and ζ_i	71
4	A Study on the Distribution of the Scaled Hankel Matrices Eigenvalues	75
4.1	Introduction	75
4.2	The effect of N	76
4.3	The patterns of $\bar{\zeta}_i$	78
4.4	The patterns of σ_i^2 and CV_i^2	79
4.5	The empirical distribution of ζ_i	80
4.6	The effect of the mean and variance of Gaussian noise distribution	87
4.7	Effect of L on the distribution of the largest and smallest eigenvalues for WN	90
4.8	Summary	93
5	Chaos and Noise	96
5.1	Introduction	96
5.2	Mathematical definition of chaos	102
5.3	Distinction between chaos and noise	104
5.3.1	Simulated data	104
5.3.2	Real data	110
5.4	Summary	112
6	Selecting the Number of Eigenvalues	113
6.1	Introduction	113
6.2	Further development on the proposed approach to the identification of r	115
6.2.1	General description	115
6.2.2	General example	118
6.2.2.1	White noise process	118
6.2.2.2	Signal plus noise	119

6.2.3	Algorithm	125
6.2.3.1	Stage 1:	125
6.2.3.2	Stage 2	126
6.3	Results	127
6.3.1	Simulated data	127
6.3.2	The effect of noise level	128
6.3.3	Selection of L	130
6.3.4	Real data	131
6.4	Summary	141
7	Application in Biomedical Data	142
7.1	Noise correction in gene expression data	142
7.1.1	Real data	144
7.1.2	Main results	145
7.2	Removing noise from EEG signal	151
7.2.1	Synthetic data analysis	152
7.2.2	Real data analysis	158
7.2.2.1	EEG data selection	158
7.2.2.2	Removing noise	159
7.2.2.3	Strange patterns	162
7.2.2.4	Discrimination of EEG signals	164
7.3	Summary	168
8	Conclusions and Future Research	170
8.1	Conclusions	170
8.2	Future Research	175
A	Linear and Non-Linear Dependencies	177
A.1	Linear correlation coefficient	177
A.1.1	Pearson correlation	177
A.1.2	Spearman correlation	178
A.1.3	Kendall correlation	178
A.2	Mutual information	179
B	Measures of Accuracy and Statistical Significance of Noise Free Time Series Reconstruction	181
B.1	Root mean square error (RMSE)	181
B.2	Mean absolute error (MAE)	182
C	Measures for Normality	183
C.1	The Shapiro-Wilk test for normality	183
C.2	D'Agostino-Pearson normality test	184

References

186

Publications

The thesis consists of work carried out by the author and the research results. He has represented most of his findings in a number of journals. Below are his publications. The co-authors are not involved in any part of the thesis. The contribution of Ghodsi M and Ghodsi Z is that they only provided the genetic data, and did proofreading.

- With Hassani, H. (2016). A new approach for selecting the number of the eigenvalues in singular spectrum analysis. *Journal of the Franklin Institute*, **353** (1), pp. 1–16. This paper is related to chapter 6.
- With Ghodsi, Z., and Hassani, H. (2016). Noise correction in gene expression data: A new approach based on subspace method. *Mathematical Methods in the Applied Sciences*, **39** (13), pp. 3750–3757. This is related to chapter 7.
- With Hassani, H., and Ghodsi, M. (2015). A study on the empirical distribution of the scaled Hankel matrix eigenvalues. *Journal of Advanced Research*, **6** (6), pp. 925–929. This paper is related to chapter 4.
- With Ghodsi, M., and Hassani, H. (2015). The empirical distribution of the singular values of a random Hankel matrix. *Fluctuation and Noise Letters*, **14** (3), pp. 1–18.
- With Hassani, H., and Ghodsi, M. (2014). Distinguishing chaos from noise: A new approach. *International Journal of Energy and Statistics*, **2** (2), pp. 137–150. This paper is related to chapter 5.
- With Hassani, H., and Ghodsi, M. (2014). A short note on the pattern of singular values of a scaled random Hankel matrix. *International Journal of Applied Mathematics*, **27** (3), pp. 237–243.
- A novel approach for removing noise from EEG signal. *Biomedical Signal Processing and Control*, **39** (2018), pp. 23–33.

List of Figures

1.1	Illustration of an ECG recording machine, adopted from [54].	8
1.2	Plot of the Lorenz attractor.	9
1.3	A Henon series (left), and Henon map (right).	11
1.4	A plot of y_t with different time delay for ECG (top) and EEG (bottom) series (see chapters 5 and 8 for more information).	11
2.1	Recording positions of 64-channel EEG data on a head. The top signifies a forehead, and projections on the right and left sides mean ears [154].	28
2.2	w -correlation of 4 EEG channels [154].	28
3.1	Histogram of ζ_i ($i = 1, 5, 10$) for matrix \mathbf{B} (right), and the histogram of λ_i ($i = 1, 5, 10$) for matrix \mathbf{XX}^T (left).	57
3.2	Plot of $\bar{\zeta}_i$ and $\bar{\lambda}_i$ ($i = 1, 10$) for different sample size N for a white noise series.	57
3.3	Plot of ζ_1 for different values of L . Example 1.	69
3.4	Plot of ζ_1 (left) and ζ_2 (right) for different values of L . Example 2.	69
3.5	Plot of ζ_1 (left) and ζ_2 (right) for different values of L . Example 3.	70
3.6	Plot of ζ_i ($i = 1, \dots, 4$) for different values of L . Example 4.	70
3.7	Plot of \bar{I}_A for different sample size N	72
3.8	Plot of $\bar{\zeta}_1$, and $\bar{\zeta}_2$ for different sample size N	74
4.1	Plot of $\bar{\zeta}_i$, ($i = 1, \dots, 10$) for different values of N	77
4.2	Plot of σ_i^2 for different values of N for WN.	78
4.3	The plot of $\bar{\zeta}_1$ (left) and σ_1^2 (right) for different values of N for WN.	78
4.4	Plot of σ_i^2	80
4.5	Plot of CV_i	81
4.6	Histograms of ζ_1 , ζ_5 , and ζ_{10} for cases ((a), ..., (d)).	82
4.7	Density of ζ_i , $i = 1, \dots, 6, 10$ for cases ((e), ..., (h)).	83
4.8	Matrix of correlation between ζ_i and ζ_j , $i, j = 1, \dots, 10$ for all cases.	85
4.9	Matrix of correlation between ζ_i and ζ_j , $i, j = 1, \dots, 10$ for the WN process considering Pearson, Kendal, and Spearman correlations.	87
4.10	Plot of $\bar{\zeta}_i$, σ_i^2 and CV_i (top), the histograms of ζ_1 , ζ_5 , and ζ_{10} (middle) and the $Q - Q$ plots (bottom) for $\mu = 0$ and $\sigma^2 = 5$	90
4.11	Plot of $\bar{\zeta}_i$, σ_i^2 and CV_i (top), histograms of ζ_1 , ζ_5 , and ζ_{10} (middle) and $Q - Q$ plots (bottom) for $\mu = 1$ and $\sigma^2 = 1$	91
4.12	Histogram of ζ_1 for different values of μ and σ^2	91

4.13	Plot of the largest eigenvalue for the white noise process with respect to different values of L	93
4.14	Plot of the smallest eigenvalue for the white noise process with respect to different values of L	94
5.1	A Henon series (left) and white noise series (right).	99
5.2	A Henon map.	100
5.3	White Noise series, map and histogram of ζ_i ($i = 1, 2$).	106
5.4	Henon series, Henon map and histogram of ζ_i ($i = 1, 2$).	106
5.5	Logistic series, Logistic map and histogram of ζ_i ($i = 1, 2$).	106
5.6	Lorenz series, Lorenz map and histogram of ζ_i ($i = 1, 2$).	107
5.7	Tent series, Tent map and histogram of ζ_i ($i = 1, 2$).	107
5.8	Matrix of correlation between ζ_i and ζ_j , $i, j = 1, \dots, 10$ for the white noise and chaotic series.	108
5.9	ECG series and histogram of ζ_i ($i = 1, 2$).	111
5.10	Plot of y_t with time delay 22 (left) and time delay 230 (right) for the ECG series.	112
6.1	Skewness coefficient of ζ_i for different values of L for the white noise series.	119
6.2	Matrix of correlation between ζ_i and ζ_j , $i, j = (1, \dots, L)$ (left) and correlation coefficient between ζ_i and ζ_{i+1} (right) for the white noise series.	120
6.3	Five sine wave series with added noise (left) and its periodogram (right).	121
6.4	Logarithm of $\bar{\zeta}_i$ for different values of L for the five sine wave series.	122
6.5	Coefficient of skewness of ζ_i for different values of L for the five sine wave series.	123
6.6	Coefficients of kurtosis and variation of ζ_i for $L = 60$ for the five sine series.	123
6.7	Correlation matrix of ζ_i for $L = 60$ (left) and the correlation coefficient between ζ_i and ζ_{i+1} (right) for the five sine series.	124
6.8	Root mean square error (RMSE) and mean absolute error (MAE) between the original signal and reconstructed series by r ($r = 1, \dots, 60$) for $L = 60$	125
6.9	Value of skewness coefficient of eigenvalues for case a, considering different values of SNR.	129
6.10	Value of skewness coefficient of eigenvalues for case c, considering different values of SNR.	129
6.11	Original (black) and reconstructed (blue) series for cases a, b, c, and d.	136
6.12	Noisy ECG series (left), the correlation matrix (middle) and the skewness coefficient (right) of ζ_i for the ECG series.	136
6.13	Histograms of the abnormal ECG (left) and normal EEG (right) series with normal curve (red line).	139
6.14	Noisy EEG series (left), the correlation matrix (middle) and the skewness coefficient (right) of ζ_i for the EEG series.	140

6.15	Original (black) and reconstructed (blue) series for ECG, and EEG series (first 100 observations).	140
7.1	Experimental data from the <i>Drosophila melanogaster</i> embryo; (a): <i>bcd</i> , (b): <i>cad</i> , (c): <i>eve</i> , (d): <i>gt</i> . y axis represents gene expression profile, which tell us what actually it is doing at a point in time (x axis).	145
7.2	Skewness coefficient (left) and the variation coefficient of ζ_i (right) for the first series of <i>bcd</i> , <i>cad</i> , <i>eve</i> and <i>gt</i> data.	147
7.3	Correlation between ζ_i and ζ_{i+1} for the first series from each data set.	148
7.4	Original (black) and extracted signal (red); (a): <i>bcd</i> , (b): <i>cad</i> , (c): <i>gt</i> , (d): <i>eve</i>	150
7.5	Improving signal extraction using sequential SSA. Original (black) and extracted signal (red);(a): <i>eve</i> , (b): <i>gt</i>	151
7.6	Rossler series (left), and Rossler + noise (right).	153
7.7	Logarithm of $\bar{\zeta}_i$ for the Rossler + Noise series.	153
7.8	Results of <i>Skew</i> , <i>Kurt</i> , <i>CV</i> , and the correlation matrix.	154
7.9	Root mean square error (RMSE) between the original Rossler signal and reconstructed components by eigentriples (left), and the reconstructed or extracted signal and the original signal series (right).	155
7.10	Typical example series, $s_t^{(1)} + s_t^{(2)}$ (left) and $y_t = s_t^{(1)} + s_t^{(2)} + \epsilon_t$ (right).	156
7.11	Logarithm of $\bar{\zeta}_i$ for the simulated series, Example 2.	156
7.12	Results of the <i>Skew</i> , <i>Kurt</i> , <i>CV</i> , and the correlation matrix for Example 2.	157
7.13	Root mean square error (RMSE) between the original signal and reconstructed components by eigentriples (left), and the plot of the reconstructed signal and original signal series (right), Example 2.	157
7.14	Specimen of the normal (left) and epileptic seizure (right) series.	159
7.15	Density of the normal and epileptic series.	159
7.16	Results of all measures corresponding to the normal and abnormal EEG series.	161
7.17	Extracted signals (top) and extracted noise (bottom) for healthy and epileptic seizure EEG.	162
7.18	Plot of y_t with time delay 10 for both reconstructed series, and with time delay 20 for the two residual series.	163
7.19	Improving signal (left) and noise (middle) extractions of epileptic EEG signal, and plot of y_t with time delay 20 for the noise series (right).	164
7.20	Results of <i>Skew</i> (ζ_i), <i>Kurt</i> (ζ_i), and <i>CV</i> (ζ_i) for normal and epileptic EEG segments.	165
7.21	Density of of the distribution of $\zeta_{1,he}$ (normal) and $\zeta_{1,ep}$ (epileptic).	166

List of Tables

4.1	Coefficient of skewness for ζ_i , ($i = 1, \dots, 10$), for all cases.	84
4.2	p -value of the D'Agostino-Pearson test for ζ_i , ($i = 1, \dots, 10$), for all cases.	84
4.3	Non-linear correlation between ζ_1 and ζ_i for all cases.	88
4.4	Average of p -values of the Shapiro-Wilk test of ζ_i for different values of μ and σ^2 of Gaussian distribution.	92
5.1	Coefficient of skewness for ζ_i , ($i = 1, 2$) for white noise and chaotic series.	106
5.2	p -value of the D-P test for ζ_i , ($i = 1, 2$) for white noise and chaotic series.	107
5.3	Mutual information $I(\zeta_1, \zeta_i)$ ($i = 2, \dots, 10$) and the standard measure for mutual information $\xi(\zeta_1, \zeta_i)$ for white noise and chaotic series.	108
5.4	p -value of the K-S test between ζ_i for the white noise series and ζ_i for chaotic series.	109
5.5	Descriptive statistics of ECG series.	111
5.6	Coefficient of skewness and p -value of the D-P test for ζ_i ($i = 1, 2$), for the ECG series.	111
6.1	Skewness coefficient of ζ_i for cases a, . . . , d.	128
6.2	Kurtosis coefficient of ζ_i for cases a, . . . , d.	128
6.3	Variation coefficient of ζ_i for cases a, . . . , d.	128
6.4	Correlation coefficient between ζ_i and ζ_{i+1} for cases a, . . . , d.	128
6.5	Root mean square error between the original and reconstructed series by ($r = 1, \dots, 12$) for cases a, . . . , d.	128
6.6	Mean absolute error between the original and reconstructed series by ($r = 1, \dots, 12$) for cases a, . . . , d.	129
6.7	Skewness coefficient of ζ_i for cases a, . . . , d and different values of L	132
6.8	Kurtosis coefficient of ζ_i for cases a, . . . , d and different values of L	132
6.9	Variation coefficient of ζ_i for cases a, . . . , d and different values of L	132
6.10	Correlation coefficient between ζ_i and ζ_{i+1} for cases a, . . . , d and considering different values of L	133
6.11	Value of w -correlations of the signal reconstruction for different values of L and r for cases a, . . . , d.	133
6.12	Simulated series: value of $Skew$, $Kurt$, CV , ρ_s , w -correlation, MAE, and RMSE of the signal reconstruction step for different values of L and r for case a.	134

6.13	Simulated series: value of <i>Skew</i> , <i>Kurt</i> , <i>CV</i> , ρ_s , w-correlation, MAE, and RMSE of the signal reconstruction step for different values of L and r for case b.	134
6.14	Simulated series: value of <i>Skew</i> , <i>Kurt</i> , <i>CV</i> , ρ_s , w-correlation, MAE, and RMSE of the signal reconstruction step for different values of L and r for case c.	135
6.15	Simulated series: value of <i>Skew</i> , <i>Kurt</i> , <i>CV</i> , ρ_s , w-correlation, MAE, and RMSE of the signal reconstruction step for different values of L and r for case d.	135
6.16	Descriptive statistics of ECG and EEG series.	139
7.1	Values of r based on <i>Skew</i> and <i>CV</i> for the ten <i>bcd</i> series.	148
7.2	Values of r based on <i>skew</i> and <i>CV</i> for the ten <i>cad</i> series.	148
7.3	Values of r based on <i>skew</i> and <i>CV</i> for the ten <i>eve</i> series.	148
7.4	Values of r based on <i>skew</i> and <i>CV</i> for the ten <i>gt</i> series.	149
7.5	Final result obtained in noise-signal separation study for gene data.	149
7.6	Descriptive statistics of for three segments; series 1, series 50 and series 100 from normal and epileptic EEG signals.	160
7.7	p -value of the D-P test for three segments; series 1, series 50 and series 100 from normal and epileptic EEG signals.	160
7.8	p -value of the D-P test for $\zeta_{1,he}$ and $\zeta_{1,ep}$	166
7.9	The p -value of KS test between ζ_i ($i = 1, 50, 100$) for the normal and epileptic EEG series.	167
7.10	Extracted features of 200 segments from the two classes.	167

Abbreviations

SVD	Singular Value Decomposition
SSA	Singular Spectrum Analysis
WN	White Noise
$N(.,.)$	Normal distribution
$U(.,.)$	Uniform distribution
Exp	Exponential distribution
$W(.,.)$	Wishart distribution
D-P	D’Agostino-Pearson test
S-W	Shapiro-Wilk test
LE	Lyapunov exponents
SLE	Stochastic Lyapunov exponent
K-S	Kolmogorov-Smirnov
$c.d.f$	Cumulative distribution function
ESD	Empirical spectral distribution
LSD	Limiting spectral distribution
RMSE	Root mean square error
MAE	Mean absolute error
RMT	Random matrix theory
\sin	Sine wave
SNR	Signal to noise ratio
LRF	Linear recurrence formula

Symbols

Y_N	Time series
N	Length of time series
L	Window length
\mathbf{I}	Identity matrix
\mathbf{X}	$L \times v$ random matrix
$\mathbf{X}\mathbf{X}^T$	The singular values of \mathbf{X}
\mathbf{H}	Hankel matrix
\mathbf{H}^T	The conjugate transpose of a Hankel matrix \mathbf{H}
\mathbf{A} or $\mathbf{A}_{L,N}$	$\mathbf{H}\mathbf{H}^T$ which is the singular values of \mathbf{H}
\mathbf{B} or $\mathbf{B}_{L,N}$	$\mathbf{A}/tr(\mathbf{A})$
\mathbf{D}	$exp(\mathbf{B})$
tr	Trace of a matrix
det	Determinant of a matrix
$I_{\mathbf{A}}$	$4 \times$ determinant of a matrix \mathbf{B}
ρ	Pearson's correlation coefficient
ρ_s	Spearman's correlation coefficient
τ	Kendall's correlation coefficient
ρ_w	Weighted correlation between two series
$I(.,.)$	Mutual information between two random variables
$\xi(.,.)$	Non-linear correlation
λ_i	The eigenvalue of the matrix $\mathbf{X}\mathbf{X}^T$
γ_i	The eigenvalue of the matrix \mathbf{A}
ζ_i or $\zeta_i^{L,N}$	The eigenvalue of the matrix \mathbf{B}

n	A natural number
m	A natural number
$\bar{\zeta}_i$	The average of ζ_i for m simulated series
σ_i^2	The variance of ζ_i
$CV(\zeta_i)$ or CV_i	The coefficient of variation of ζ_i
$Skew(\zeta_i)$	The coefficient of skewness of ζ_i
$Kurt(\zeta_i)$	The coefficient of kurtosis of ζ_i
$\rho_s(\zeta_i, \zeta_j)$	Spearman's coefficient between ζ_i and ζ_j
μ	Mean of Gaussian noise distribution
σ^2	Variance of Gaussian noise distribution
\mathbf{G}	$L \times L$ positive definite matrix
v	A positive integer for the number of degrees of freedom
$\ \cdot \ _F$	Frobenius norm
η	Vector of diagonal elements in $\mathbf{\Lambda}$
φ	Vector of diagonal elements in $\mathbf{\Gamma}$
$ \rho_s $	The absolute value of ρ_s
U_i	Vector
V_i	Vector
$(\sqrt{\gamma_i}, U_i, V_i)$	i -th eigentriple of the SVD
ϖ_v	Random spectral measure
ϖ	Deterministic measure
\mathbb{R}^L	Euclidean space of dimension L
\mathbb{H}^L	L -trajectory space of a time series Y_N
\mathcal{H}	Hankelisation operator
\mathcal{L}_r	r -dimensional linear subspace of \mathbb{R}^L

Acknowledgements

I would like to express sincere appreciation and gratitude to my supervisors Dr. Hossein Hassani and Professor Steve Letza, for their invaluable assistance and guidance throughout my research. Dr. Hassani's significant knowledge, creativity, and contribution in the field of Singular Spectrum Analysis have encouraged me to do my research on this topic. Furthermore, the main parts of this thesis came about as a result of several published studies. I would like to thank the co-authors Dr. Hossein Hassani and Dr. Mansi Ghodsi for their supervision, encouragement and useful suggestions.

I would also like to show my gratitude to King Saud bin Abdul-Aziz University for Health Sciences for the financial support they have given me throughout this research. My thanks also go to all my colleagues and staff of the Business School at Bournemouth University for their kindness and help during my research.

My deep gratitude goes to all my family, especially my parents, who have given me so much in my life. I am really grateful to them and they will always be my role models.

I dedicate this thesis to my parents and my beloved wife Afaf, whose unconditional encouragement and support made it possible for me to commence my PhD studies. Without Afaf's love, patience and support, this work would never have been completed. I wish to express my heartfelt love to my sons Yazan and Yamin for coping with the undue paternal deprivation during my research.

Declaration of Authorship

I, Nader Alharbi, declare that this thesis, entitled ‘Separability between Signal and Noise Components Using the Distribution of Scaled Hankel Matrix Eigenvalues with Application in Biomedical Signals’, and the work presented in it are my own. I confirm that:

- This work was conducted for the degree of a Doctor of Philosophy at Bournemouth University.
- Where any part of this thesis has previously been submitted or published, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- I have acknowledged all main sources of help.

Signed:

Date:

Chapter 1

Introduction

1.1 Overview of singular spectrum analysis and separability

Time series analysis is an important subject in various areas of science and engineering. The subject plays a significant role, for example, in system identification, prediction and pattern classification. However, the observed time series or the signal is often nonlinear and contaminated by noise [1]–[4], which needed to be removed or reduced for a proper analysis. Thus, separability in time series analysis is an initial and important task, which can be done by decomposing the original time series into sums of series, so that each component in the sum can be identified as either a main component (signal) or noise [5]. In this regard, the concept of separability between signal and noise components in time series characterises how well those different components can be separated from each other. In practice, exact separability of time series components rarely happens, but approximate separability can very often be attained [6]. Therefore, for analysing and filtering noisy real-life time series we can only suppose approximate separability to produce less noisy series.

The issue of filtering time series to be less noisy with minimum loss of information has been widely studied; see, for example, [5], [7]–[9]. There exist several approaches which can be implemented in order to reduce noise from measured time series data

and to extract the signal which we are interested in. For example, the traditional method to reduce noise is to use the linear filter [1]. However, it is suggested that such a method is not suitable in some cases; for example, in the case of discontinuously varying data [1]. Furthermore, it is difficult to separate the noise from the signal with traditional methods, which may distort the signal and make the situation even worse [10]. There are several nonlinear noise reduction methods such as simple nonlinear filtering [11], local average [12], local projective [13], chaos-based smoothing [2, 14], wavelet thresholding [15, 16], adaptive wavelet thresholding [17], adaptive filtering and wavelet shrinkage [18], the structural function method [19], the factorised Hankel optimization method [20], and singular value decomposition (SVD) [5], [21]–[23].

Recently, it has been accepted that the singular spectrum analysis (SSA) technique, which is based on SVD, can be used as an alternative to traditional digital filtering methods [5]. For example, it has been shown that the SSA technique outperforms other traditional methods in filtering biomedical data [24]. Furthermore, the effect of noise reduction on the linear and nonlinear measure of dependencies was considered in [25]. The results indicate that the SSA-based technique can be used as a noise reduction method for filtering either noisy financial or chaotic series. The efficiency of noise reduction for curve fitting in nonlinear growth curve models was also studied in [5]. The results confirm that SSA can be used as a powerful tool for noise reduction in longitudinal measurements. Moreover, in [26], the SVD based approach was used for clustering and outlier detection in biomedical data. In addition, [27] suggested a new approach for the separation of murmurs from heart sound, based on SSA.

The appearance of SSA is usually associated with publication of the papers [28, 29] in 1986. There are five books dedicated to SSA [6],[30]–[33]. Two books, [30] and [31], only provide an elementary introduction to SSA. The book [32] develops the main methodological principles, and considers several theoretical issues. Its aim is to establish SSA as an important technique for time series analysis, which indeed helped to attract the attention of researchers from different scientific communities to SSA. [6] was published in 2013, is fully devoted to the methodology of SSA and

also includes some new topics. [33] was recently published in late December 2015, and is the latest book that focuses on SSA, an effective approach for single channel biomedical signal analysis, and its bivariate, multivariate, tensor-based, complex-valued, quaternion-valued and robust variants. Furthermore, SSA as a powerful approach in time series analysis, has been developed and applied to many practical problems; see, for example, [34]–[36] and references therein.

The SSA approach is a non-parametric technique that combines elements of classical time series analysis, multivariate statistics, multivariate geometry, dynamical systems and signal processing. It decomposes the original time series into a sum of small numbers of interpretable components, such as slowly varying trend, oscillatory component and noise. SSA is based on the use of the singular value decomposition approach for calculating eigenvalues and eigenvectors of the trajectory or Hankel matrix \mathbf{H} of the data. Note that the random Hankel matrix is one of the main matrices that play a significant role in various fields of mathematics and statistics [37]. It naturally appears in multivariate analysis and signal processing, particularly in SSA, context where \mathbf{H} is a data matrix, the column vectors of which represent the L -lagged vector of observations in \mathbb{R}^L [27, 38].

The SSA technique consists of two complementary stages: decomposition (embedding and SVD), and reconstruction (grouping and diagonal averaging). A short description of the SSA technique as follows (more details are given in Chapter 3). Consider a one-dimensional series $Y_N = (y_1, \dots, y_N)$ of length N . Transferring this series into the multi-dimensional series H_1, \dots, H_K with vectors $H_i = (y_i, \dots, y_{i+L-1})^T \in \mathbb{R}^L$ provides the following trajectory matrix:

$$\mathbf{H} = [H_1, \dots, H_K] = (h_{i,j})_{i,j=1}^{L,K} = \begin{pmatrix} y_1 & y_2 & y_3 & \cdots & y_K \\ y_2 & y_3 & y_4 & \cdots & y_{K+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_L & y_{L+1} & y_{L+2} & \cdots & y_N \end{pmatrix}, \quad (1.1)$$

where L ($2 \leq L \leq N - 1$) is the window length and $K = N - L + 1$. Obviously, $h_{ij} = y_{i+j-1}$, so the trajectory matrix \mathbf{H} is a Hankel matrix, which means all the elements along the diagonal $i + j = \text{const}$ are equal. Define $\mathbf{H}\mathbf{H}^T$, where \mathbf{H}^T is the conjugate transpose. Denote by γ_i ($i = 1, \dots, L$) the eigenvalues of $\mathbf{H}\mathbf{H}^T$ taken in decreasing order of magnitude ($\gamma_1 \geq \dots \geq \gamma_L \geq 0$) and by U_1, \dots, U_L the orthonormal system of the eigenvectors of the matrix $\mathbf{H}\mathbf{H}^T$ corresponding to these eigenvalues. The square root of the eigenvalues of the matrix $\mathbf{H}\mathbf{H}^T$ are called singular values of \mathbf{H} . Set

$$d = \text{rank } \mathbf{H} = \max(i, \text{ such that } \gamma_i > 0),$$

note that in real-life series we usually have $d = L^*$ with $L^* = \min\{L, K\}$.

The SVD of the trajectory matrix can be written as:

$$\mathbf{H} = \mathbf{H}_1 + \dots + \mathbf{H}_d, \quad (1.2)$$

where $\mathbf{H}_i = U_i \sqrt{\gamma_i} V_i^T$, U_i is the eigenvector corresponding to the eigenvalue γ_i ($i = 1, \dots, L$) of $\mathbf{H}\mathbf{H}^T$, and V_i is the principle component defined as $V_i = \mathbf{H}^T U_i / \sqrt{\gamma_i}$. The collection $(\sqrt{\gamma_i}, U_i, V_i)$ is called the i -th eigentriple of the SVD. The grouping procedure partitions the set of indices $\{1, \dots, d\}$ into m disjoint subsets I_1, \dots, I_m . Let $I = \{i_1, \dots, i_p\}$. Then the resultant matrix \mathbf{H}_I corresponding to the group I is defined as $\mathbf{H}_I = \mathbf{H}_{i_1} + \dots + \mathbf{H}_{i_p}$. The resultant matrices are calculated for groups $I = I_1, \dots, I_m$ and the expansion (1.2) can be rewritten as:

$$\mathbf{H}_I = \mathbf{H}_{I_1} + \dots + \mathbf{H}_{I_m}. \quad (1.3)$$

The way of selecting the sets is called eigentriple grouping. In the final step of analysis we use the diagonal averaging procedure to transform each matrix \mathbf{H}_{I_j} to the form of a Hankel matrix, which can be subsequently transformed into a new time series. The reconstructed series produced by the elementary grouping will be called elementary reconstructed series.

The description of SSA above can be summarised in two main stages: decomposition and reconstruction. These stages depend upon two main selections; namely, window length L and the number of required eigenvalues/singular values, denoted by r , for reconstruction. It is mentioned in [6] that for optimal values of L , the weighted correlation or w -correlations (a common measure for approximate separability) between two reconstructed series is approximately zero. Furthermore, since the signal components are often dominating, the number r of the leading components related to the signal can be estimated using the matrix of w -correlations between the elementary reconstructed components [6]. Thus, an appropriate choice of L and r leads to a perfect analysis and separability between time series components.

Auxiliary information about the original time series is always helpful in selecting L and r [6, 32]. For example, if it is known that the time series has a periodic component with integer period, then it is advisable to take L proportional to that period. Furthermore, there are several rules for the selection of L that have a practical basis (see section. 2.4.3 [6]). It was discussed in [30] that for a series of length N , selecting $L = N/4$ is common practice. It was also recommended that L should be large enough but not larger than half of the series [32]. In 2010, Golyandina [39] provided general recommendations on the selection of L to attain minimal error, and the paper demonstrated that optimal selection depends on the particular problem. It was shown that the error behaviour depends on the type of noise, and the selection of the window length to one-half of N was approved as suitable in most cases. A minimum description length (MDL) criterion was presented in [40], which can be applied to automatically choose both the window length and the signal. The authors showed that under general conditions the criterion will determine the true signal dimension with probability one as the sample size increases, and will select the smallest L consistent with the embedding theorem. The application of the rule $L = \beta N$ was examined by [39] and it was concluded that the use of a value of β close to 0.5 will provide optimal separation and reconstruction.

In 2012, Hassani et al. [41] found that for a series of length N and the optimal choice of the number of the eigenvalues r for reconstructing the signal, the suitable value of L is the median of $\{1, \dots, N\}$. Moreover, [42] established numerical bounds on the

mean squared reconstruction error and presented asymptotic bounds for the mean squared separation error to show how optimal evaluation for the window length can be made practically. According to the literature, previous studies only consider L -separability; however, the results of [41] indicate that we also need to consider r for optimal separability. Despite various attempts that have been applied, there is no universal rule for obtaining optimal selections of L and r . In this thesis, we are interested in the signal as a whole; separability of signal components is not important to us and is not our goal. Golyandina and Zhigljavsky [6] point out that for signal extraction as a whole the importance of the selection of the value of L is diminished. For example, for a signal including a periodic component, divisibility of L by its period is not essential for separating the whole signal from the noise. All these motivations led us to focus on the study of r -separability. Although our focus is on the optimality of selection of the value of r , we also briefly consider the best choice of the value of the window length.

Accordingly, the goal is to determine the accurate dimension of the system, which is the smallest dimension with which the filtered series is reconstructed from a noisy signal. In this case, the main analysis is based on the study of the eigenvalues and corresponding eigenvectors. If the signal component dominates the noise component, then the singular values of the Hankel matrix have a few large singular values and many small ones, suggesting that the variations of the data take place mainly in the eigenspace corresponding to these few large singular values. A similar situation happens if the length of the series is large.

Although SSA has been applied widely in the analysis of various data from different disciplines, its use in biomedical signals has received little attention. SSA applications in biomedical signals, include: extraction of a weak fetal heart signal from a noisy maternal electrocardiogram (ECG) [34], separation of biomedical data such as Electromyography (EMG) affected by an ECG signal [43]; image processing [44]; DNA microarray gene expressions [35]; infant electroencephalography (EEG) seizure detection [45]; and analysing sleep EEG data [46].

Here we develop the SSA approach for analysing and filtering biomedical signals. The task in analysing biomedical signals is to remove noise and extract important information from biomedical data. However, analysing biomedical signals is not an easy task, as its data is often nonstationary, particularly when the signal includes an abnormal event [47]. Although several methods have been applied for analysing and discriminating between different categories of biomedical signals [47], many of them are based on restrictive assumptions of normality, stationarity and linearity of the observed data. An improved version of the SSA technique is used here, as it is applicable for analysing nonlinear and nonstationary time series, which could be robust for analysing and modelling biomedical data.

1.2 Biomedical signals and chaos

Biomedical signals are observations that carry significant information about the physiological activities of organisms and embrace many subjects of interest, ranging from gene and protein sequences [48], to neuromuscular control [49], neural and cardiac rhythms [50], tissue and organ images [51, 52], and sensory perception, control and coordination [53].

The human body is considered as a complex system which is affected by its environment, but is able to adapt itself to any change in life. To understand such a system, structure and the functions of the human body, which help in the diagnosis, treatment and monitoring of diseases, and in making decisions, we need to study the biomedical measurements, which can be in the form of time series that reflect functions such as heart, brain, blood pressure, muscles, respiration, eye pressure, skin impedance, body temperature, enzymes, and many other biometrics. An example of an electrocardiogram (ECG) recording machine is illustrated in Figure 1.1, which prints a record of our heartbeat onto a paper strip.

Statisticians and engineers are proposing new approaches to the analysis of these signals, using various mathematical formulae and algorithms. These signals can be calculated using different software and medical devices to provide doctors with

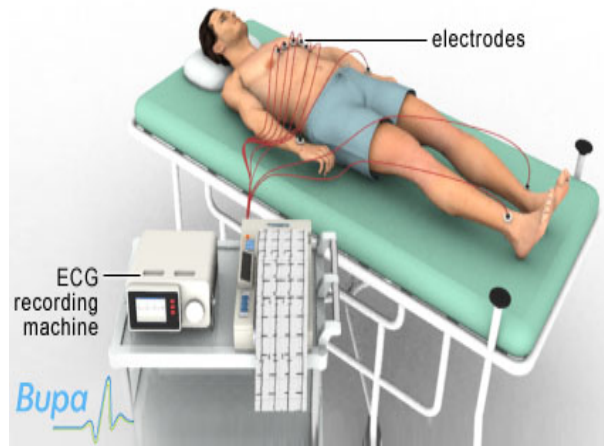


FIGURE 1.1: Illustration of an ECG recording machine, adopted from [54].

real time data and significant insights to aid in clinical evaluations. Real time observations can lead to proper assessment and management of chronic diseases, and earlier identification of negative events such as heart attacks, epileptic seizures and strokes. Thus, the observations are used and analysed for earlier diagnosis of different diseases. Alzheimer's disease, for instance, is one of the main diseases that can be diagnosed in a timely fashion using biomedical signal processing. Physicians combine EEG records with other testing parameters to attempt to detect patterns that can distinguish Alzheimer's patients from people who have other mental illnesses. In this case, biomedical signal processing is mainly beneficial in the serious care setting, where patients' data has to be analysed in real time.

Biomedical signals of interest are often corrupted by noise due to the randomness of the world around them, which makes them complex signals. This also hides and distorts the necessary information involved in the signal. In this case, it is vital to understand these signals and how to obtain the important information.

Furthermore, biomedical signals are different in their nature. Measured biomedical data includes different types of noise and nonlinear trends which require separation before any statistical evaluation of the signal can be carried out. For example, EEG and EMG recordings contain sum of semi sinusoid components, noise, and also other related signals. To identify such signal sources, separation and noise removal have been widely used. In many cases, the connectivity of the sources and the trend of the signal rhythms is subject to change. For EMG sources, for instance, the challenging

task is to detect actively connected potentials from noisy signals. The separation of such sources and classification of these signals have also been extensively studied. However, the methods used do not mimic the original data features and therefore are not efficient [33]. Many of the physiological measurements have a small number of channels; the SSA technique can analyse the data to determine the frequency signature of the signal over time. In addition, SSA is a subspace approach, since cases can be well explained by a small number of distinct orthogonal bases.

The proposed approach for SSA is not limited to any assumption such as the stationarity of the data. The ability to identify eigenvalues corresponding to noise components is a novel feature of the proposed approach presented in this thesis. This provides a certain confidence in the identification of the biomedical signal subspace.

As mentioned above, the human body system is a complex system. Chaos theory is a complicated mathematical theory that can be applied to studying complex dynamical systems. Some dynamical systems, such as logistic maps are chaotic everywhere, but in several situations chaos can be found only in a subset of phase space. The most interesting cases appear when the chaotic behaviour takes place on an attractor. The attractor that occurs from chaotic systems is known as a strange attractor. The first and most known complex attractor is Lorenz attractor [55], which gives a very interesting and strange pattern which looks like the wings of a butterfly (see Figure. 1.2).

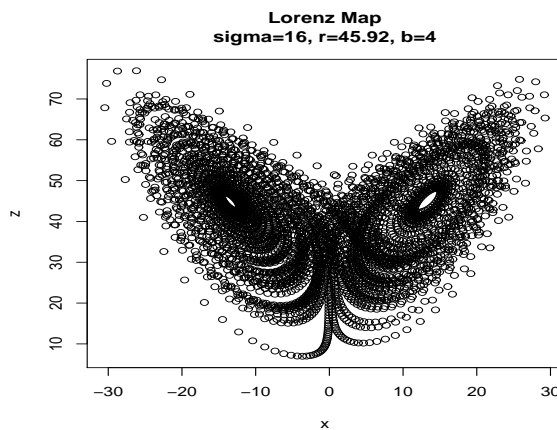


FIGURE 1.2: Plot of the Lorenz attractor.

An EEG signal, for example, is considered as a nonlinear time series [2, 56], particularly during an epileptic seizure, and can also be characterised as being chaotic [57]. Chaotic behaviour can be found in many nonlinear dynamical systems in nature [56]. Therefore, there has been critical attempts to differentiate strange attractors in brain signals [58]. Furthermore, as decomposing brain signals is a necessary tool for detecting chaotic behaviour or epileptic activity, the extracted information from EEG recordings or detection of epileptic seizure is helpful for diagnosing and treating epilepsy patients.

During biomedical signal problem processing, chaos theory can provide a potential explanation for the various, erratic, and complex patterns that arise in most biomedical signals [59]. Generally, several applications of nonlinear techniques are applied to biomedical problems, which span from heartbeat studies [60]–[63], to brain systems [64]–[66], blood pressure regulation [67, 68], nervous and muscles systems [69]–[73], breathing rhythms [74], cardio-respiratory coordination [75], genomic, deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) sequences [75, 76], and complex physiological systems of human and animal bodies [77, 78].

As stated earlier, observed time series often contain noise which needs to be removed or reduced for better analysis. Noise reduction for chaotic time series has also been widely studied; see, for example, [1], [79]–[81]. There are several noise reduction methods suitable for chaotic time series; see [82]–[85] and references therein. Many of those methods are based on statistical and dynamical aspects which use embedding methods such as SVD method [28] and delays [86]. The SVD-based method is considered as a very effective approach to noise reduction from chaotic signals [87, 88]. However, the main problem is that the deterministic system, chaos, and stochastic process, noise, have several common properties that can make them undistinguishable [89]. The distinction between them is considered as one of the most important topics; see, for example, [90]–[96]. Figure. 1.3 (left) illustrates an example of a chaotic Henon series, which looks random, but transferring this series to a two dimensional map gives us an attractive pattern (see Figure. 1.3 (right)). The proposed approach provides a valuable solution to this problem. Figure. 1.4

shows an example of our results for analysing real ECG and EEG data; details can be found in Chapters 5 and 7.

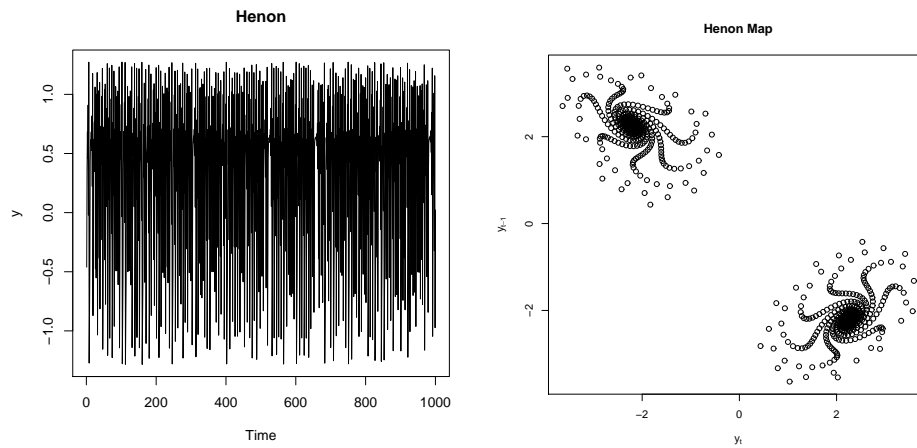


FIGURE 1.3: A Henon series (left), and Henon map (right).

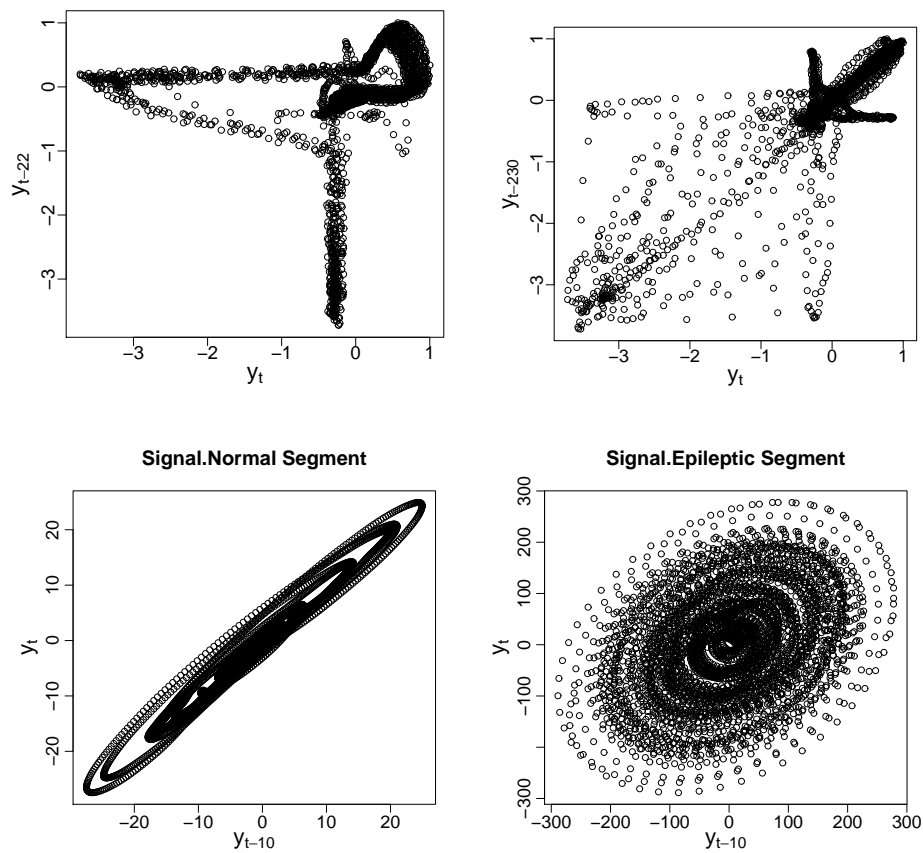


FIGURE 1.4: A plot of y_t with different time delay for ECG (top) and EEG (bottom) series (see chapters 5 and 8 for more information).

1.3 Aims, objectives, contributions and organisation of the thesis

The main aim of this research is to introduce a new approach to the selection of the optimal value of r for SSA, which can make a novel contribution to the area of noise reduction and filtering in signal processing and time series analysis. The proposed approach is mainly used to identify the required number of eigenvalues corresponding to the signal component, which relies on the distribution of the scaled Hankel matrix eigenvalues. The approach does not need any assumption; for example, stationarity of the series or linearity of the residual. This shows that the proposed approach can be considered as a promising one for the extraction of any biomedical signal contaminated in linear, nonlinear, stationary, or non-stationary noise, or buried in other signals.

First, to achieve the objectives of the thesis, the distribution of the scaled Hankel matrix eigenvalues generated from symmetric and nonsymmetric distributions with various criteria are studied in depth. Furthermore, we present the theoretical and methodological aspects of the proposed approach, and demonstrate that it can be considered as an alternative one for time series analysis and forecasting. Second, the eigenvalue distribution is considered as a novel approach to the differentiation between and identification of chaos from noise. Third, since the main issue in extracting signals using the SSA procedure lies in identifying the number of eigenvalues needed for signal reconstruction, the proposed approach has been developed to select an optimal value of r . In addition, we also consider the selection of the value of the window length.

This thesis also seeks to explore the applicability of the proposed approach to eigenvalue identification in several simulated and real biomedical signals with different structures and signal to noise ratios. Another objective is to apply the approach to eigenvalue identification to four different gene expression protein profiles. In addition, the performance of the approach in filtering EEG signals, extracting strange

patterns, and discriminating between normal and abnormal EEG signals is discussed.

The proposed approach and results of the thesis make a novel contribution to:

- The selection of the optimal value of r for the separability between signal and noise components in SSA, with applications in biomedical and genetic time series.
- Improving the quality of the reconstruction step in SSA for signal extraction.
- The area of distinction between deterministic and stochastic processes.
- The application area of classification of brain signals.

The thesis is organised into eight chapters, which can be summarised as follows:

- Chapter 1: An overview of the problems to be addressed in this thesis and presentation of the aims and objectives of the work. We also provide a brief introduction to biomedical signals and chaos, decomposition these signals and extraction of the sources of interest.
- Chapter 2: This chapter includes a general overview of applications in biomedical signals and genetic studies with SSA.
- Chapter 3: The theoretical aspects of singular spectrum analysis and a brief mathematical background are presented. This chapter includes new theoretical results on the eigenvalues of $\mathbf{A} = \mathbf{H}\mathbf{H}^T$ divided by its trace, with new proposals on the eigenvalues, proofs of their properties, the effect of window length and rank of Hankel matrix on the eigenvalues, and the asymptotical behaviour of the determinant and eigenvalues of the scaled Hankel matrix. The chapter also includes several examples related to the theoretical results obtained, particularly the effect of window length and rank of the Hankel matrix on the eigenvalues.

- Chapter 4: A comprehensive study on the distribution of the scaled Hankel matrix eigenvalues using simulated data, including symmetric and nonsymmetric distributions, trend series, and sine waves. The chapter also includes the effect of the sample size or the length of the series on the eigenvalues, the patterns of the mean and variance, the coefficient of the variations for each eigenvalue, the empirical distribution for each eigenvalue, the effect of the mean and variance of Gaussian distribution on the eigenvalues, and some examples related to the effect of the window length on the distribution of the largest and smallest eigenvalues for the white noise process.
- Chapter 5: The distribution of Hankel matrix eigenvalues is considered as a novel approach for distinguishing chaos from noise. The applicability of the proposed approach is examined using ECG data.
- Chapter 6: The chapter includes the main results of the thesis together with the theoretical results in chapter 3 and the study in chapter 4. The principle of the algorithm for selecting the optimal value of r is presented. The algorithm is applied to a number of simulated and real data with different structures and signal to noise ratio. Furthermore, the best choice of L is briefly studied.
- Chapter 7: This chapter provides successful applications of the proposed approach. The approach is applied to remove noise from *Drosophila* segmentation genes. Four different types of gene data are used, analysed and evaluated by using different criteria. In addition, the chapter presents another application of the approach. It includes analysis of EEG signals, removing noise, discrimination between EEG signals and extraction of strange patterns. Several criteria based on the largest eigenvalue are introduced and used as features to distinguish between normal and epileptic EEG segments.
- Chapter 8: Presents the general conclusions of the thesis and proposes possible improvements and directions for future research work.
- Finally, Appendix A includes the linear and nonlinear measures of dependence between two series, Appendix B presents a description of the measures of

accuracy and statistical significance of noise free time series reconstruction, and Appendix C gives a description of the measures for normality.

Chapter 2

Literature Review of SSA

2.1 Literature review

Singular spectrum analysis (SSA) is a technique of time series analysis and forecasting. It combines elements of classical time series analysis, multivariate statistics, multivariate geometry, dynamical systems and signal processing. The aim of the SSA technique is to decompose the main series into a sum of a small number of interpretable components such as a slowly varying trend, oscillatory components and a structureless noise. It is based on the singular value decomposition (SVD) of a specific matrix constructed upon the time series. Neither a parametric model nor stationarity-type conditions have to be assumed for the time series, which makes it a model-free method and hence allows this method to have a broad range of applicability.

The basic SSA method includes two complementary stages: decomposition and reconstruction; each one contains two steps. At the first stage the series is decomposed and at the second one the original series is reconstructed, then it is used for forecasting purpose. The general concept that we study the properties of SSA is separability, which characterizes how well we can separate mixed components from each other. The approximate separability is often absent in the observed series that has a complex structure. For series with complex and special structures, there are

various ways of adapting SSA leading to diverse versions such as SSA with single and double centering, sequential SSA and Toeplitz SSA [32]. In addition, two versions of SSA, based on the perturbation theory and minimum variance estimator, were proposed in [97].

Furthermore, SSA can be used for analyzing relatively short series, which is a very important feature of this method. It has been illustrated that SSA works very well for both short and long time series in forecasting macro-economics data [98]. Note that despite some probabilistic and statistical concepts are used in the SSA-based methods, no statistical assumptions need to be made such as normality of the residuals and stationarity of the series.

Additionally, SSA is a capable tool which can be applied for solving many problems such as; smoothing, finding trends of different resolution, simultaneous extraction of cycles with small and large periods, extraction of seasonality components, extraction of periodicities with varying amplitudes and simultaneous extraction of complex trends and periodicities.

The commencement of the SSA method is generally associated with publication in 1986 of the papers [28, 29] by Broomhead and King. Since the publication of that papers SSA has received a very good attention in literature. In addition to [28, 29] the list of most cited papers on SSA published in the 1980s and 1990s includes [84], [99]–[101].

There are five books dedicated to SSA [6],[30]–[33]. Two books, [30] and [31], give a primary introduction to SSA. The book [32] develops the main methodological principles, and considers several theoretical issues. The aim of the book is to introduce SSA as a very powerful technique for time series analysis, which really aided to attract the attention of researchers from different scientific communities to SSA. [6] was published in 2013, is fully devoted to the methodology of SSA and contains other different topics. [33] was recently published in late December 2015, and is the latest book that focuses on SSA, an effective approach for single channel biomedical signal analysis, and its bivariate, multivariate, tensor-based, complex-valued, quaternion-valued and robust variants. During the last 14 years the SSA method

has been developed and new successful applications of it across different fields have been reported; see for example, [34]–[36] and references therein.

Links between SSA and other methods of time series analysis. The SSA technique is not linked with GARCH, ARIMA, wavelets and other methods of this type. However, it has close links with few methods of multivariate statistics and with methods of signal such as projection pursuit and principal component analysis (PCA); see [6]

Theory of SSA. We usually refer to the book [32] for the basic theory of SSA. Since 2001, a number of effective studies on theoretical aspects of SSA have been published. The important theoretical monograph on perturbations in SSA and subspace-based methods of signal processing is [102]. There is also an important theoretical paper that is [103], where the separability concept is further studied and developed for the signal plus noise model. [39] also discussed the elements of the theory of SSA.

SSA for change-point detection and subspace tracking. Assume that the points y_1, y_2, \dots , of the series arrive sequentially in time and we use basic SSA for the series at hand. Then the distances can be monitored from the sequence of the trajectory matrices to the r -dimensional subspaces, and the distances between these r -dimensional subspaces can also be constructed. Any significant change in these distances may show a change in the mechanism generating the time series. It is worth mentioning that this change in the mechanism does not have to influence the whole series structure but rather only a few of its components. For more details; see [32] (chapter 3), and [104].

Monte-Carlo SSA. A general assumption of SSA about the noise in the signal+noise model is the association of noise with a series that cannot be well approximated by a time series of finite rank. If the noise is assumed to be stochastic and red then the so-called Monte Carlo SSA is a common technique to be applied. This version has special tests, which are based on the Monte Carlo simulations and are devised for evaluating the hypothesis of the presence of a weak signal on the background of a large noise, see [100].

SSA for density estimation. It was shown in [105] that SSA can be applied for the estimation of nonparametric density and also could give estimates that are more precise than the celebrated Kernel density estimates.

SSA for multivariate time series. Multivariate SSA (MSSA) [30, 31] is an extension of the basic SSA for simultaneous analysis of different time series. This method could be useful for decomposing various series with common structure. It could also be applied for introducing a causality between two time series. In [6], it is mentioned that the absence of Granger [106] causality of two series $Y_N^{(1)}$ and $Y_N^{(2)}$ means that the information of $Y_N^{(1)}$ does not improve the quality of forecasts of $Y_N^{(2)}$. The discussion of MSSA causality is shown in [107].

2D-SSA for image processing. 2D-SSA is an extension version of both basic and multivariate SSA for decomposing images. There is only one difference between these three versions, which is in the construction of the Hankel matrix, see [31, 44, 108]. However, it should be noted that the moving window in 2D-SSA is a rectangle and the window length becomes a result of two numbers. This means that the the Hankel matrix could be large and a very good implementation of the SVD becomes important.

Comparison of SSA with other techniques. [39, 102] compare the SSA technique with some subspace-based techniques of signal processing. Most of these techniques are based on the assumption that the original series is random and stationary; they include some techniques that are famous in signal processing, such as Karhunen-Loeve decomposition (for signal processing references see, for example [36]). It has been accepted that SSA, which is based on SVD, can be used as an alternative to traditional digital filtering methods [29]. It has been shown that SSA is better than others such as smoothing method, digital butterworth filters, and splines in filtering biomedical signals [109]. Numerical comparison of SSA with ARIMA, ARAR algorithm, Holt-Winter and other classical methods of time series analysis can be found in many papers of the volume [110] and in several papers devoted to applications of SSA, see for example [35, 36, 107, 111–113]. Moreover, the SSA method has also been used for signal extraction and forecasting of the UK tourism

income. The results indicated that SSA is better than SARIMA and time-varying parameter State Space models in terms of several forecasting accuracy criteria [32].

Application areas. It has been proven that SSA is a very useful and used as a standard method in the analysis of climatic, meteorological and geophysical time series; see, for example, [99, 101] (climatology), [114] (meteorology), [115] (marine science), [116] (geophysics); for more researches, see [30–32, 84, 99–101, 116, 117] and the studies in [110]. There are recent areas of application of SSA including engineering, image processing, medicine, actuarial science and many other fields; see, for example, [111, 118, 119] and different papers in [110]. Econometrics and medicine are special cases as SSA was fundamentally unknown only a few years ago in these two fields, however, it has recently made a tremendous advancement and has become more popular; see, for example, [34, 43, 46, 107, 113, 120].

As stated above, the description of SSA above can be summarised in two main stages: decomposition and reconstruction. These stages depend upon two main selections; namely, window length L and the number of required eigenvalues/singular values, denoted by r , for reconstruction. It is mentioned in [6] that for optimal values of L , the weighted correlation or w -correlations (a common measure for approximate separability) between two reconstructed series is approximately zero. Furthermore, since the signal components are often dominating, the number r of the leading components related to the signal can be estimated using the matrix of w -correlations between the elementary reconstructed components [6]. Thus, an appropriate choice of L and r leads to a perfect analysis and separability between time series components.

It is always useful using the auxiliary information about the original time series for the choice of the values of L and r [6, 32]. For instance, if we know that the original time series has a periodic component with integer period, then we should take L proportional to that period. In addition, there are different ways for the selection of L that have a practical basis (see section. 2.4.3 [6]). It was suggested in [30] that for a series of length N , the choice of $L = N/4$ is an appropriate and common in practice. [32] recommend that L should be large enough but not larger

than half of the series. Golyandina [39] recommends some ways for the selection of L to attain minimal error, and the results showed that the optimality of the selection is based on the main problem. The results also showed that the residual behaviour depends on the noise type, and the selection of the window length to one-half of N was shown to be suitable in most cases. [40] present a minimum description length (MDL) criterion, which can be applied to automatically choose both the window length and the signal. The results showed that under general conditions the criterion will determine the true signal dimension with probability one as the sample size increases, and will select the smallest L consistent with the embedding theorem. The application of the rule $L = \beta N$ was examined by [39], and the author points out that the use of a value of β close to 0.5 will give optimality in the separation and reconstruction.

Recently, Hassani et al. [41] suggest that the appropriate value of L is the median of $\{1, \dots, N\}$ for reconstruction. In addition, [42] show numerical bounds on the mean squared reconstruction error and present asymptotic bounds for the mean squared separation error to indicate how the evaluation for the selection of the window length can be made in practice.

Based on the literature, previous researches only focus on L -separability; but, [41] show that the value of r should be considered for good separability. Although several studies that have been done, there is no general rule for selecting L and r . In this thesis, the interest is in the signal as a whole; separability of signal components is not our aim. Golyandina and Zhigljavsky [6] point out that for signal extraction as a whole the importance of the selection of the value of L is diminished. For instance, if the signal contains a periodic component, divisibility of L by its period is not important for extracting the whole signal. Thus, we focus here on the study of r -separability. Despite we mainly focus on the optimality of selection of the value of r , the best choice of the value of the window length will be considered briefly.

Here we develop the SSA in the selection of the values of r , then apply it for analysing and filtering biomedical signals. As we mentioned earlier, the applications of SSA in biomedical data has received little attention despite it has been

widely used in different disciplines. In the following subsections, we give a review of publications in biomedical and genetic studies with SSA.

2.2 A Review of Applications in Biomedical Signals and Genetic Studies with SSA

2.2.1 Applications in biomedical signals with SSA

In biomedical processing, the important feature of the SSA technique is that it can be used for analysing biomedical signals and separate these signals from each other and also remove unwanted component of noise. For example, in the decomposition of EEG records, it is important to obtain neurophysiologically useful information in applications such as seizure detection, brain computer interfacing (BCI), and sleep analysis, which can be done by using SSA [46].

For multichannel biomedical data, use of the blind source separation (BSS) approach is traditional and can help in separating the mixed signals into their constituent sources [121]–[123]. Blind source separation (BSS), also known as blind signal separation, is the separation of a set of source signals from a set of mixed signals, without the help of information about the source signals or the mixing process. There are different methods of blind signal separation such as principal components analysis, independent component analysis, singular value decomposition, dependent component analysis and stationary subspace analysis. However, BSS fails to analyse single channel recordings [46]. There are many applications where only one channel is used; for instance, restoring EMG mixed with ECG artifact [124, 125]; separating noise from the neuronal spikes, which is considered as a hard task [126]; several BCI applications; and various sleep stages in a recorded channel [127, 128]. The desired information in such applications can be only retrieved from special channels. On the other hand, the use of BSS for separability in recordings with a small number of channels is often weak [46].

Among the many methods used, SSA is a relatively new technique that has been proven to be a successful approach. Recently, SSA has been used for analysing biomedical signals and has shown its strong potential for such studies [34]. Below we give a review of publications in biomedical studies with SSA.

In 2004, SSA was applied to backscattered ultrasound signals from in vitro human cancellous bone specimens [129]. The authors explained that there are many approaches, such as temporal and spectral autocorrelation, higher order statistics, power spectrum and cepstrum, and quadratic transformation, which can be used for mean scatterer spacing (MSS) estimation; this is used to detect variations in quasiperiodic tissue microstructures. However, those approaches characterise signals that include only a mixture of periodic and nonperiodic structures, while SSA is often used in nonlinear dynamics to mainly identify components of signals related to periodic contributions and then identify dominant acyclicity. Thus, SSA can be more effective than other methods for separability between periodic, nonperiodic and noise components. The results showed that applying SSA to backscattered ultrasound signals may be beneficial for presenting information related to tissue microarchitecture that is not obvious in clinical images.

Moreover, SSA was used in [109] for the smoothing of biomechanical signals, particularly kinematic ones. Four examples were presented in this study to illustrate the capability of SSA smoothing and show its superiority over other techniques used in biomechanical analysis such as digital butter worth filter, splines and filter based on spectral analysis. The authors also showed that SSA works effectively on both stationary and nonstationary time series. They mentioned that its only drawback was that there are no general rules for selecting window length; however, it was shown in their examples that the results were similar for different values of window length. The authors recommended that in some cases using sequential SSA can be better than the basic SSA for better separability between signal trend and noise.

Furthermore, EEG recordings can include artifacts created by ocular activity, especially those signals recorded from frontal channels. The blinking or movements of eyes produce an electrical signal or artifact called the electrooculogram (EOG).

This artifact (EOG) can appear in the recorded EEG signals, which need to be removed. In 2005, a method based on SSA was presented in [130] for removing EOG from EEG signals. In the study, the feature vectors were clustered and the principle components (PCs) were calculated locally within each cluster. Note that in pattern recognition and machine learning, a feature vector is an n -dimensional vector of numerical features that represents some objects. Subsequently, the EOG signal was assumed to be related to the PCs corresponding to the largest eigenvalues. The minimum description length (MDL) criterion; which is a formula in which the best hypothesis for a given set of data is the one that leads to the best compression of the data, was applied to determine the number of eigenvectors required to represent the EOG artifact. Thus, the authors first extracted the EOG artifact and then deducted it from the original EEG signal to attain the corrected EEG signal. The MDL criterion was selected to be used for the EOG artifact removal as it yielded the most consistent results compared to the results of the variance criterion and Akaike's information criterion (AIC). Note that given a collection of models of the data, AIC estimates the quality of each model and provides a means of model collection. The results also illustrated that a development method of SSA by combining a clustering step works very well with EEG data where EOG signals display large amplitudes.

In some recent published researches, such as [34] and [131]–[134], the choice of periodic components has been considered by clustering the eigentriples [131], or by proposing different criteria [132]–[134]. Furthermore, Ghodsi et al. [26], for example, present a new approach based on SSA for the detection of temporomandibular disorders (TMDs) which appear as a result of issues with the jaw, temporomandibular joint (TMJ); and facial muscles, which are responsible for chewing and moving the jaw [135, 136]. In this approach, the movement data of markers located on points of special interest on the faces of many cases was extracted and analysed. The participants were grouped into a set of healthy subjects and one of individuals with the temporomandibular disorder in order to extract the components corresponding to the signal and to separate the noise. The results verified that the SSA method can be used as a detection and noise reduction method for analysing biomedical signals

with abnormal behaviour, such as the data of temporomandibular disorder analysis. The outcomes also illustrated that the extracted noise can be used for detection. In particular, a helpful pattern was extracted for classification of the volunteers with TMD and healthy participants using the noise series. Thus, it was proposed that the extracted noise includes useful information which can be considered for detection.

Fetal heart rate (FHR) and variability of its beats are of the most useful indicators of the health of the fetus [34]. The record of the maternal electrocardiogram signal is a mixture of mother's and the fetal heart signals, and it is often very noisy. The fetal electrocardiogram (FECG) provides information about FHR and its health. For example, when we have information about FHR, we can detect any arrhythmia or variations in the fetal heart rate. Furthermore, the heart waveform includes helpful diagnostic information. For instance, the position and the sign of each part in the cardiac waveform is very important to distinguish between different forms of supraventricular tachycardia. Several researches have been devoted to fetal electrocardiogram extraction. For example, the method based on SVD [22, 23]; neural networks [137]; cross-correlation techniques [138]; adaptive filtering combined with genetic algorithms [139]; adaptive filtering [140, 141]; decomposition into an orthogonal basis [142]; fuzzy logic [143]; independent component analysis for blind source separation [144, 145]; the frequency based technique [146]; real-time signal processing [147]; wavelet based techniques [148]–[150]; and principal component analysis and projective filtering [151]. However, some of the mentioned studies have many major drawbacks [34].

In 2010, Ghodsi et al. [34] proposed a multivariate singular spectrum analysis (MSSA) technique for separating and extracting the mother's electrocardiogram (MECG) signal, the fetal electrocardiogram (FECG) signal and the noise component from the mixed ECGs. In fact, MSSA was used to capture the fetal heart signal, which is a classical problem in biomedical engineering. A significant feature of MSSA for the combined signal is that; in MSSA the dynamic structure between the FECG and MECG signals can be captured in order to extract the FECG signal, since the SSA method consists of the elements of multivariate geometry, multivariate

statistics and dynamical systems. The results confirmed the capability of the SSA technique for extraction of the FECG signal and removal of the noise component. The results also illustrated that SSA can be easily adapted to a broad class of biomedical signals.

Recordings of respiratory sounds are often mixed with heart sound interference. In some of the heart sound cancellation techniques, the important preprocessing task is to localise essential components of the heart sound. SSA was used in [134] to localise heart sounds in respiratory signals. The results of the SSA technique were compared with some techniques that use the wavelet transform and entropy (ENT) of the signal to detect the components of the heart sound. Error in localisation and the correlation between the primary heart sound and outputs of the techniques used were applied to assess the performance of the techniques on synthetic data. False positive and negative detections were considered as a measurement tool that can be applied to both synthetic and real data. The results indicated that SSA was significantly better than the wavelet-based method for false detection and correlation with the underlying heart sounds. Furthermore, efficiency of the SSA technique was slightly better than that of the ENT method. In addition, the execution time of the SSA method was much lower than that of the entropy method.

Adaptive line enhancer (ALE) is a method that is used as an effective filter for filtering measured cyclic signals from contaminated Gaussian white noise. However, it depends fundamentally on second order similarity between the signal and its delayed form, and is more valuable and successful for the narrowband signal [43]. In 2012, Sanei et al. [43] proposed a new ALE based on SSA. In the SSA-based ALE, the eigentriples were adaptively selected for the reconstruction stage using the delayed form of the data. In that research, both the main ALE and SSA-based ALE were examined against various noise levels. The results illustrated that the proposed algorithm based on SSA is more effective than the main ALE. The authors mentioned that the approach can be used for the signals with some periodic components. However, biomedical signals do not often have a precise period and their periods are different; such signals are called quasiperiodic. They overcame this issue by providing some solutions. The most simple one by selection of shorter

segments, within which the signals remain periodic. A more comprehensive solution is by warping the cycle intervals around average for each segment of the signal. The results demonstrated that the technique based on SSA is powerful for the extraction of several physiological signals mixed with stationary, or nonstationary noise, or burned in other periodic or non-periodic signals. The results also showed that the technique enhances the quality of surface electromyography. In addition, the approach was able to identify the EMG noise from the ECG signal.

Heart and lung sounds are recorded from approximately the same region of the human body. It is difficult to diagnose the lung sound due to the involved heart sound. Thus, the separation of the two sounds is a required step for a precise diagnosis. In 2013, the SSA-based ALE method was also used for separation of the heart sound artifact from respiratory signals [152]. This method is a new noise reduction technique that has the feature of adaptive selection of needed eigentriples in the reconstruction stage of SSA. In the study, the method was evaluated using synthetically mixed and real respiratory signals. The result of the SSA-based ALE method was compared with the traditional ALE technique. The performance of the SSA-based ALE method outperformed the traditional ALE one with respect to the output signal to noise ratio and correlation coefficient.

The study by [153] is an attempt to show the value of the SSA method in the analysis of the cardiac RR time series data of a healthy, congestive heart failure, and atrial fibrillation subjects. The author used SSA and the Monte Carlo SSA (MCSSA) test to identify the noise component existing in the series. Furthermore, SSA was applied in [154] to an exponential-cosine sequence and measured 64-channel EEG data (see Figure. 2.1). The authors showed that SSA is able to extract important information from the data series. Figure. 2.2 illustrates the results of w -correlation. They also showed that the use of SSA can identify the characteristics of EEG data. In addition, they discussed that it is possible to find the optimal parameters in the sense of the forecast by applying SSA forecasting performed by using the linear recurrence formula (LRF).

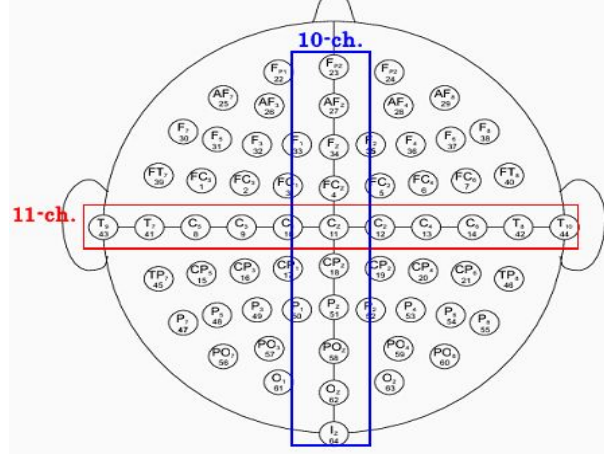


FIGURE 2.1: Recording positions of 64-channel EEG data on a head. The top signifies a forehead, and projections on the right and left sides mean ears [154].

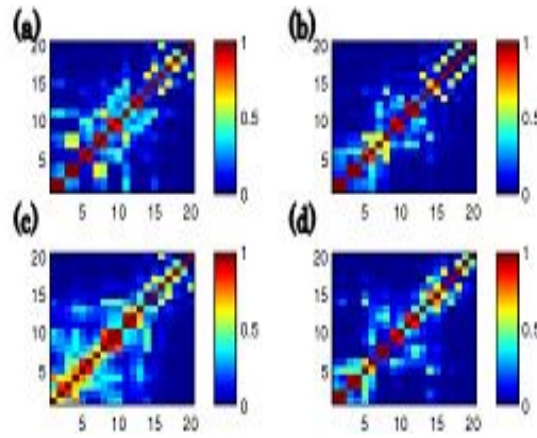


FIGURE 2.2: w -correlation of 4 EEG channels [154].

Murmur is an indicator of diverse heart abnormalities. In 2011, Sanei et al. [27] suggested an adaptive singular spectrum analysis (ASSA) technique as a technique for separability between murmur and heart sound (HS). In that article, SSA was adapted to the variations in the statistical properties of the data and successfully applied to detect murmur from heart sound. The results indicated a precise separability of normal heart sound from the murmur segment. Furthermore, a comparison between ASSA and the other general methods such as neural networks, sup-optimal transform and PCA, clearly illustrates that ASSA is better in terms of accuracy of the results and the overall computation speed or cost [27]. Those methods ignored the overlap and the correlation between the two sounds. The authors pointed out that one of the main features of ASSA is that it effectively separates the heart and

murmur signals even over the temporally overlapped regions, which is indicated by the low values of the correlation in most types of murmurs. They also pointed out that clinicians can use those separated signals for diagnosis of other heart problems.

In 2015, the performance of the basic SSA algorithm was improved in [46] by using tensor decomposition instead of the traditional SVD method for decomposition of the EEG single channel signal. The frequency diversity of the data was used to determine the subspace of interest. In that study, sleep EEG data, and the stages of sleep for the subjects in normal condition, with sleep extension, and with sleep restriction, were precisely estimated and compared with the results of sleep scoring by clinical experts. The results showed that the method can help in analysis and understanding of sleep EEG data.

2.2.2 Applications in genetic studies with SSA

In recent years, SSA has been accepted as a valuable technique in quantitative genetic studies. Presented here is a review of recent publications in genetic research with SSA. Since SVD is a main stage of SSA, some SVD applications in gene expression data are also presented briefly. According to the review, the most important areas of using SSA in genetics are denoising and signal extraction, which indicates that SSA can be used as a reliable and promising approach for genetic analysis.

Nowadays, there are a huge number of datasets in the area of genetics and expression measurement, and there exist several diverse approaches and methods for decomposing genetic datasets [155]–[157]. Validity of expression measurements are no longer accepted as the main issues in working with these types of data, but without using proper decomposition tools, the accuracy of the inferences from the data is hard to understand [158]. If the study results are achieved from an inadequate model, the result can not be interpreted in a scientific format [159].

Parametric methods have been applied historically for analysis of such data [160, 161]. However, the restricting pre-assumptions required for parametric methods reduces their use and increases the use of nonparametric techniques. In recent

years, it has been widely accepted that nonparametric approaches can be used as an alternative method for decomposing genetic data due to their innate nature [6], and thus the applications in genetic and biomedical areas have been developed and expanded. The SSA method, as a nonparametric method, is considered to be very successful approach; it has been recently applied to genetic data and has shown its capability in such studies [162, 163].

The capabilities of signal extraction and filtering, derivation of an analytical formula of the signal, together with batch processing of a set of similar series, are the major features of the SSA technique in the field of genetics [164]. It is in researchers' interest to use SSA for signal extraction in gene expression data and filtering in microarrays. The reason behind their main interest in the SSA technique is its potential for noise reduction, and because of the fact that genetic data is often mixed with considerable noise. Denoising of such data is considered as a hard task when analysing genetic data [165, 166].

Several researches have been conducted to obtain quantitative data in genetics using the microarray method. The microarray is a capable method for analysing large amounts of genes simultaneously, which makes it a very useful. However, microarray data are often contaminated with a significant level of noise, which can affect the performance of the statistical results and make extraction of signals a very challenging task [167].

In the following part we detail the available applications of the SSA technique for noise reduction and signal extraction in genetics, which can also be found in [168].

In 2006, the first application of SSA was reported; its use was for signal extraction from *Drosophila melanogaster's* gene expression profile [164]. Similar study was then used in [169], showing an improved result. A study of signal extraction from the noisy *Bicoid (Bcd)* protein profile in *Drosophila melanogaster* was presented in [117]. The issue under study was complex for two reasons; the data are extremely noisy and include outliers, and the structure of the noise is unknown. In this study, a small window length was used with addition of a constant to the original series to improve separability and reconstruct the signal. Furthermore, SSA was used for

reduction of the two types of noise; the noise was detected because of variability in the nuclear order and experimental noise [170].

The SSA technique based on minimum variance was proposed in [97], which was then used in [118] for filtering and extracting the *bcd* gene expression signal. The results indicated that SSA based on minimum variance is able to outperform the other methods used for denoising *Bcd* [118].

Furthermore, one of the important applications of the microarray method is the study of rhythmic cellular processes, which can exhibit cyclic behaviours [168, 171]. Consequently, one of the main tasks in gene expression data analysis is to detect subset of genes that display cyclicity in their gene expression time series profiles. However, the time series of gene expression are commonly of short length, with very few durations and highly contaminated with noise [171]–[173]. Various gene products regulate those rhythmic cellular processes, which can be measured by using multiple deoxyribonucleic acid (DNA) microarray experiments. Time series expression data corresponding to the rhythmic behaviour of a specific gene can be obtained from a group of gene experiments over a time period.

In 2008, a new procedure was proposed in [162] for analysing the cyclicity of gene expression profiles, and extracting the trend using SSA and the Autoregressive (AR) model. The useful combination of SSA and the AR model enabled the authors to extract more periodic genes than in the use of the classical Fourier analysis technique in [174]. Subsequently, the procedure was used and developed in four research papers [35], [163], [175, 176], and showed its capability and success in detecting periodicity from 60% to 80%. For more information about the combination of SSA and AR for detecting periodic profiles, see also [177].

Despite SVD being a step in the SSA technique, it has been applied independently as an important and useful technique for analysing microarray data; see, for example, [178]. The authors considered the elimination of cross-hybridization in real-time microarray data. The authors also proposed techniques such as SVD for separating the components of the composite signal in order to obtain an estimation of the amounts of hybridising and cross-hybridising targets.

Although the topology of gene networks can be obtained from measurements of variation in gene expression over time, measurement devices such as microarrays yield information on thousands of genes over few biological replicates, which are complicated to analyse. In 2010, an iterative empirical Bayesian procedure and algorithm were presented in [179] for inferring the structure of gene networks, and the SVD of the Hankel matrix was used for selecting the model. This method allowed them to minimise the needed computation time for running the algorithm, as it reduced the requirement to run it over a huge number of values for the hidden state dimension. Furthermore, it was discussed in [180] that the use of SVD for analysing microarray data is not time extensive. Moreover, in this study, SVD was applied to a Hankel matrix of gene expression data, and the singular values corresponding to the large magnitude were used as an estimate for the most correct state space dimension. In particular, the singular values of the estimated Hankel autocovariance matrix were computed and standardised to a 0 – 1 scale. Those singular values of magnitude larger than the threshold were performed as an estimation for the state space dimension.

In addition, two dimensional SSA (2D-SSA) has been considered for decomposing two-dimensional scalar fields [6]. In the 2D-SSA approach, it is necessary to select two different values of the window length, whereas in univariate SSA only one window length is needed (for more information see [108]). In 2012, the 2D-SSA approach was used for measuring between the nucleus variability or noise observed in the gradient of *Bcd* in *Drosophila* embryos [181].

According to the literature discussed above, SSA is a reliable and flexible technique for signal extraction and gene expression modelling. The studies of the extraction of the signal from the segmentation gene profile in *Drosophila* embryos and the studies of the rhythmic behaviour of a particular gene recommend that this approach can be of general use in assessing other expressional systems. Therefore, this technique can be helpful in analysing any noisy data [168]. However, as we already mentioned that the main issue in extracting signals using the SSA method lies in identifying the optimal values of L and r needed for signal reconstruction. According to the literature, despite several attempts that have been applied, there is no general rule

for selecting the optimal values of L and r . Furthermore, most of these attempts are based in some assumptions. In this thesis, the proposed approach is totally different than others, and can be applied universally without any assumptions.

2.3 Summary

According to the literature, previous studies only consider L -separability; however, the results of [41] indicate that we also need to consider r for optimal separability. Despite various successful attempts that have been applied, there is no universal rule for obtaining optimal selections of L and r . In this thesis, we are interested in the signal as a whole; separability of signal components is not important to us and is not our goal. It was pointed in [6] that for signal extraction as a whole the importance of the selection of the value of L becomes less. For example, for a signal including a periodic component, divisibly of L by its period is not essential for separating the whole signal from the noise. All these motivations led us to focus on the study of r -separability. Although our focus is on the optimality of selection of the value of r , we also briefly consider the best choice of the value of the window length. The proposed approach has several advantages in terms of both the quality of the results and computational speed. The main advantage of the proposed approach compared to other approaches for the selection of the values of L and r in SSA, is that our approach can be used universally without any assumption. To conclude this chapter, the main aim of this thesis is to develop the SSA method then apply it to biomedical signals.

Chapter 3

Theoretical Framework

3.1 White noise

While signals can have various structures and forms, noise is often like white noise (WN) [6]. A white noise series is often part of time series in the form of an error, an unpredictable randomness. It is a weak stationary process which has a zero mean and is uncorrelated over time [182]. If the white noise process is normally distributed it is called a Gaussian white noise process [183]. White noise is important because more complicated stochastic processes are generally defined and modeled in terms of white noise [183]. Therefore, Gaussian white noise is considered as the type of noise in our simulation study.

A white noise series $Y_N = (y_1, \dots, y_N)$, where:

$$y_t \sim \text{WN}(0, \sigma^2), \quad (3.1)$$

has the following characteristics for every $t, s \in N$. Thus, y_t is a WN process $\forall t \in N$:

1. $E[y_t] = 0$.
2. $E[y_t^2] = \sigma^2$.

3. $E[y_t y_s] = 0$, for $s \neq t$,

where E is the expectation operator [182].

3.2 Singular spectrum analysis

The SSA technique consists of two main stages: decomposition and reconstruction. Each stage includes two separate steps:

- Decomposition: Embedding and SVD.
- Reconstruction: Grouping and Diagonal Averaging.

In the first stage we decompose the series and in the second we reconstruct the original series and use the reconstructed series (which is without noise) to forecast new data points. Below we provide a brief discussion of the methodology of the SSA technique (for more information see [32], chapters 1 and 2, and [6], chapter 2).

3.2.1 Embedding

The role of this step is to transfer a one-dimensional time series $Y_N = (y_1, \dots, y_N)$ into the multi-dimensional series H_1, \dots, H_K with vectors $H_i = (y_i, \dots, y_{i+L-1})^T \in \mathbb{R}^L$, where $K = N - L + 1$. Vectors H_i are called *lagged vectors*. The *window length* L is the parameter that defines the embedding procedure, which is an integer such that $1 < L < N$. The result of the embedding step is the trajectory matrix defined in (1.1), which is $\mathbf{H} = [H_1, \dots, H_K] = (h_{ij})_{i,j=1}^{L,K}$. It should be noted that the (1.1) process is a clear symmetry property: the transposed matrix \mathbf{H}^T is the trajectory matrix of the same series Y_N with window length K instead of L .

The embedding procedure is a standard technique in time series analysis, the analysis of nonlinear dynamical systems and signal processing. In SSA, the window length L should be sufficiently large. In particular, this is necessary so that each

L -lagged vector includes an important part of the structure of the initial series. Large values of L allow us to consider each L -lagged vector \mathbf{H}_j as a separate series and investigate the dynamics of certain characteristics for this collection of series. For more information and discussion on the choice of L we refer to [6], section 2.4.3.

3.2.2 Singular value decomposition

Singular value decomposition is a method that decomposes real or complex matrices into sub-matrices, which have numerous applications in signal processing and statistics. It is closely related to eigendecomposition. The eigendecomposition of the matrix $\mathbf{H}\mathbf{H}^T$ or the SVD of the matrix \mathbf{H} yields a collection of L eigenvalues and eigenvectors. We start with the general properties of the SVD, which are essential for SSA.

It has already been mentioned that the SVD of an arbitrary nonzero $L \times K$ matrix \mathbf{H} is a decomposition of \mathbf{H} in the form

$$\mathbf{H} = \sum_{i=1}^d \sqrt{\gamma_i} U_i V_i^T, \quad (3.2)$$

where γ_i ($i = 1 \dots, d$) are the eigenvalues of the matrix $\mathbf{A} = \mathbf{H}\mathbf{H}^T$ arranged in decreasing order, $d = \text{rank } \mathbf{H} = \max(i, \text{ such that } \gamma_i > 0)$, $\{U_1, \dots, U_d\}$ is the corresponding orthonormal system of the eigenvectors of the matrix \mathbf{A} and $V_i = \mathbf{H}^T U_i \sqrt{\gamma_i}$. Note that we call $\sqrt{\gamma_i}$ the singular value, while U_i and V_i are the left and right singular vectors of the matrix \mathbf{H} , respectively. If \mathbf{H}_i is defined as $\mathbf{H}_i = \sqrt{\gamma_i} U_i V_i^T$, then Eq. (3.2) can be rewritten in the form of Eq. (1.2), i.e. as the representation of \mathbf{H} as a sum of the elementary matrices \mathbf{H}_i . If all the eigenvalues have multiplicity one, then the expansion (1.2) is uniquely defined.

The expansion (3.2) demonstrates that the SVD possesses the following symmetrical property: V_1, \dots, V_d form an orthonormal system of eigenvectors for the matrix $\mathbf{H}^T \mathbf{H}$ corresponding to the same eigenvalues γ_i . The columns and rows of \mathbf{H} are

subseries of the original time series. Consequently, the left and right singular vectors also have a temporal structure and hence can also be regarded as time series.

The SVD in the expansion (1.2) is optimal in the sense that among all the matrices $\mathbf{H}^{(r)}$ of rank $r < d$, the matrix $\sum_{i=1}^r \mathbf{H}_i$ gives the best approximation to the trajectory matrix \mathbf{H} , so that $\|\mathbf{H} - \mathbf{H}^{(r)}\|_F$ is minimum. Note that the Frobenius norm of a matrix \mathbf{Z} is $\|\mathbf{Z}\|_F = \sqrt{\langle \mathbf{Z}, \mathbf{Z} \rangle_F}$ where the inner product of two matrices $\mathbf{Z} = \{z_{ij}\}_{i,j=1}^{s,q}$ and $\mathbf{Y} = \{y_{ij}\}_{i,j=1}^{s,q}$ is defined as $\langle \mathbf{Z}, \mathbf{Y} \rangle_F = \sum_{i,j=1}^{q,s} z_{ij}y_{ij}$. Note that $\|\mathbf{H}\|_F^2 = \sum_{i=1}^d \gamma_i$ and $\|\mathbf{H}_i\|_F^2 = \gamma_i$ for $i = 1, \dots, d$. Thus, the ratio $\gamma_i / \sum_{i=1}^d \gamma_i$ can be considered as the characteristic of the contribution of the matrix \mathbf{H}_i to expansion (1.2). Hence, $\sum_{i=1}^r \gamma_i / \sum_{i=1}^d \gamma_i$ the sum of the first r ratios, is the characteristic of the best approximation of the trajectory matrix by the matrices of rank r . Note that for a fixed value of L and a series with length N , the trace of matrix $\mathbf{A}_{L,N} = \mathbf{H}\mathbf{H}^T$, $tr(\mathbf{A}_{L,N})$, is as follows:

$$T_{\mathbf{H}}^{L,N} = \|\mathbf{H}\|_F^2 = tr(\mathbf{A}_{L,N}) = \sum_{i=1}^L \gamma_i. \quad (3.3)$$

Consider now the trajectory matrix \mathbf{H} as a sequence of L -lagged vectors. Denote by $\mathbb{H}^{(L)} \subset \mathbb{R}^{(L)}$ the linear space spanned by the vectors H_1, \dots, H_K . We shall call this space the trajectory space of the series Y_N . The Eq. (3.2) demonstrates that $\mathcal{U} = (U_1, \dots, U_d)$ is an orthonormal basis in the d -dimensional trajectory space $\mathbb{H}^{(L)}$.

Introducing $Y_i = \sqrt{\gamma_i}V_i$, we can rewrite the expansion (3.2) in the form $\mathbf{H} = \sum_{i=1}^d U_i Y_i^T$, and for lagged vectors $H_j = \sum_{i=1}^d y_{ji}U_i$, where the y_{ji} are the components of the vector Y_i . This means that vector Y_i is composed of the i -th components of lagged vectors H_j represented in the basis \mathcal{U} . Consider now the transposed trajectory matrix \mathbf{H}^T . Setting $Z_i = \sqrt{\gamma_i}U_i$ we yield the expansion $\mathbf{H}^T = \sum_{i=1}^d V_i Z_i^T$, which corresponds to the representation of the sequence of K -lagged vectors in the orthonormal basis V_1, \dots, V_d . Consequently, the SVD presents two dual geometric descriptions of the matrix \mathbf{H} .

The feature considered above may be reformulated in the language of multivariate geometry for the L -lagged vectors as follows. Assuming $r < d$, then among all r -dimensional subspaces \mathcal{L}_r of \mathbb{R}^L , the subspace spanned by U_1, \dots, U_r approximates these vectors in the best way, which is obtained at $\mathcal{L}_r = \text{span}(U_1, \dots, U_r)$. The ratio $\sum_{i=1}^r \gamma_i / \sum_{i=1}^d \gamma_i$ is the characteristic of the best r -dimensional approximation of the lagged vectors.

Another optimal characteristic of the SVD corresponds to the properties of the directions determined by the eigenvectors U_1, \dots, U_d . Specifically, U_1 determines the direction such that the variation of the projections of the lagged vectors into this direction is maximum. Every subsequent eigenvector determines the direction that is orthogonal to all previous directions, and the variation of the projection of the lagged vectors onto this direction is also maximum. Therefore, it is natural to call the direction of the i -th eigenvector U_i the i -th principal direction. Note that the elementary matrices $\mathbf{H}_i = U_i Y_i^T$ are built up from the projections of the lagged vectors onto the i -th particular directions.

The view above on the SVD of \mathbf{H} composed of L -lagged vectors and an appeal to association with the principal component analysis lead us to the following terminology. We shall call vector U_i the i -th eigenvector, vector V_i the i -th factor vector and vector $Y_i = \sqrt{\gamma_i} V_i$ the i -th principal component, respectively.

3.2.3 Grouping

The grouping step corresponds to splitting the elementary matrices \mathbf{H}_i in (1.2) into several groups, as in (1.3), and adding up the matrices within each group. Let us assume that $m = 2$ in (1.3), $I_1 = I = \{i_1, \dots, i_r\}$, and $I_2 = \{1, \dots, d\} \setminus I$, where $1 \leq i_1 < \dots < i_r \leq d$.

The purpose of this step is to separate the additive components of the time series. We discuss here the very interesting concept of separability. Assume that the time series Y_N is a sum of two time series $Y_N^{(1)}$ and $Y_N^{(2)}$; that is, $y_i = y_i^{(1)} + y_i^{(2)}$ for

$i = 1, \dots, N$. Assume also that L is fixed and denote by \mathbf{H} , $\mathbf{H}^{(1)}$ and $\mathbf{H}^{(2)}$, the L -trajectory matrices of the series Y_N , $Y_N^{(1)}$ and $Y_N^{(2)}$, respectively.

Consider now an SVD (1.2) of the trajectory matrix \mathbf{H} . If there is a collection of indices $I \subset \{1, \dots, d\}$, such that $\mathbf{H}^{(1)} = \sum_{i \in I}$ and $\mathbf{H}^{(2)} = \sum_{i \notin I}$, we can say that the series $Y_N^{(1)}$ and $Y_N^{(2)}$ are separable by the decomposition (1.2). In the study of separability between two matrices $\mathbf{H}^{(1)}$ and $\mathbf{H}^{(2)}$ in the expansion $\mathbf{H} = \mathbf{H}^{(1)} + \mathbf{H}^{(2)}$, the contribution of $\mathbf{H}^{(1)}$ is measured by the share of the corresponding eigenvalues: $\sum_{i \in I} \gamma_i / \sum_{i=1}^d \gamma_i$.

We can look at the separation of the series by the decomposition (1.2) from different perspectives. Suppose the set of indices is fixed $I = I_1$, and \mathbf{H}_{I_1} is the corresponding resultant matrix. If \mathbf{H}_{I_1} and \mathbf{H}_{I_2} , where $\mathbf{H}_{I_2} = \mathbf{H} - \mathbf{H}_{I_1}$, are Hankel matrices, then they are necessarily the trajectory matrices of certain time series that are separable by the expansion (1.2).

Furthermore, if the two matrices \mathbf{H}_{I_1} and \mathbf{H}_{I_2} are approximate Hankel matrices, then there exist series such that $Y_N = Y_N^{(1)} + Y_N^{(2)}$ and the trajectory matrices of these series are close to \mathbf{H}_{I_1} and \mathbf{H}_{I_2} , respectively; we can say that the series are approximately separable.

Accordingly, the goal of the grouping step is to find the groups I_1, \dots, I_m such that the matrices $\mathbf{H}_{I_1}, \dots, \mathbf{H}_{I_m}$ satisfy expansion (1.3) and are close to certain Hankel matrices.

Let us now move to the grouping step from the view point of multivariate geometry. Consider \mathbf{H} as the trajectory matrix of a time series Y_N , $Y_N = Y_N^{(1)} + Y_N^{(2)}$, and the series $Y_N^{(1)}$ and $Y_N^{(2)}$ are separable by the decomposition (1.2), which coincides with splitting the index set $\{1, \dots, d\}$ into I and $\{1, \dots, d\} \setminus I$.

Assuming $m = 2$ in expansion (1.3), this means that U_1, \dots, U_d , the basis in the trajectory space $\mathbb{H}^{(L)}$, is split into two groups of basis vectors. This coincides with the representation of $\mathbb{H}^{(L)}$ as a product of two orthogonal subspaces (eigenspaces) $\mathbb{H}^{(L,1)} = \text{span}(U_i, i \in I)$ and $\mathbb{H}^{(L,2)} = \text{span}(U_i, i \notin I)$, spanned by $U_i, i \in I$, and $U_i, i \notin I$, respectively.

Separability of $Y_N^{(1)}$ and $Y_N^{(2)}$ means that matrix \mathbf{H}_I , whose columns are the projections of the lagged vectors H_1, \dots, H_K onto the eigenspace $\mathbb{H}^{(L,1)}$, is precisely the trajectory matrix of the series $Y_N^{(1)}$.

The grouping procedure is based on the analysis of the eigenvectors U_i and V_i , and eigenvalues γ_i in the SVD expansion. The principles and methods of identifying the SVD components for their inclusion into different groups are described in [6], section 2.4. Note that each matrix of the SVD is exactly determined by the corresponding eigentriple, so we shall talk about grouping of the eigentriples rather than that of the elementary matrices \mathbf{H}_i .

It should be noted that the case of two series components ($m = 2$) considered above is often regarded as the issue of separating out a single component rather than that of separating two terms. This means only one group of indices, namely I , is interesting to us.

In signal processing, the series $Y_N^{(1)}$ is considered as a signal. Therefore, we often select $I_1 = \{1, \dots, r\}$ for some r and call $\mathbb{H}^{(1)}$ the signal subspace.

3.2.4 Diagonal averaging

The idea of using diagonal averaging or the Hankelization procedure is to convert a matrix into the form of a Hankel matrix which can be transferred to a time series. If z_{ij} stands for an element of a matrix \mathbf{Z} , then the n -th term of the resulting series is obtained by averaging z_{ij} over all i, j such that $i + j = n + 1$. The result of the Hankelization of a matrix \mathbf{Z} is the Hankel matrix $\mathcal{H}\mathbf{Z}$, which is the trajectory matrix corresponding to the series obtained as a result of the diagonal averaging. Note that Hankelization is an optimal procedure in the sense that the matrix $\mathcal{H}\mathbf{Z}$ is the nearest to \mathbf{Z} (with respect to the matrix norm) among all the Hankel matrices of the corresponding size (for more information see [32], chapter. 6, section. 2). In turn, the Hankel matrix $\mathcal{H}\mathbf{Z}$ uniquely defines the series by relating the value in the diagonals to the values in the series. By applying the Hankelization procedure to

all matrix components in Eq. 1.3, we obtain this expansion:

$$\mathbf{H} = \tilde{\mathbf{H}}_{I_1} + \dots + \tilde{\mathbf{H}}_{I_m} \quad (3.4)$$

where $\tilde{\mathbf{H}}_{I_1} = \mathcal{H}\mathbf{H}_{I_1}$. This is equivalent to the decomposition of the initial series $Y_N = (y_1, \dots, y_N)$ into a sum of m series:

$$y_n = \sum_{k=1}^m \tilde{y}_n^{(k)} \quad (3.5)$$

where $\tilde{Y}_N^{(k)} = (\tilde{y}_1^{(k)}, \dots, \tilde{y}_N^{(k)})$ corresponds to matrix \mathbf{H}_k .

Note that if \tilde{z}_{ij} is the (i,j) -th entry of the estimated matrix $\tilde{\mathbf{Z}}$, then applying the diagonal averaging formula it follows that:

$$\tilde{y}_j = \frac{1}{s_2 - s_1 + 1} \sum_{i=s_1}^{s_2} \tilde{z}_{i,j+1-i}, \quad (3.6)$$

where $s_1 = \max\{1, j+1-K\}$, $s_2 = \min\{L, j\}$ and \tilde{y}_j is the j -th of the reconstructed series \tilde{Y}_N .

Properties of the Hankel Matrix

The behaviour of $T_{\mathbf{H}}^{L,N}$ defined in Eq. (3.3) with respect to different values of L is considered in the following theorem.

Theorem 3.1. *Consider the trajectory matrix \mathbf{H} . Then,*

$$T_{\mathbf{H}}^{L,N} = \sum_{j=1}^N w_j^{L,N} y_j^2,$$

where

$$w_j^{L,N} = \min\{ \min\{L, K\}, j, K + L - j \} = w_j^{K,N}.$$

Proof. Using the definition of \mathbf{H} (1.1), we have;

$$T_{\mathbf{H}}^{L,N} = \sum_{i=1}^L \sum_{j=i}^{N-L+i} y_j^2.$$

Changing the order of the double summation above gives the following equation:

$$T_{\mathbf{H}}^{L,N} = \sum_{j=1}^N C_{j,L,N} y_j^2,$$

where $C_{j,L,N} = \min\{j, L\} - \max\{1, j - N + L\} + 1$. Therefore, we only need to show that $C_{j,L,N} = w_j^{L,N}$ for all j and L . We consider both cases, $L \leq K$ and $L > K$. For $L \leq K$, we obtain

$$C_{j,L,N} = \begin{cases} j & \text{for } 1 \leq j \leq L, \\ L & \text{for } L + 1 \leq j \leq K, \\ N - j + 1 & \text{for } K + 1 \leq j \leq N, \end{cases}$$

which is exactly equal to $w_j^{L,N}$. Similarly, for $L > K$, we have

$$C_{j,L,N} = \begin{cases} j & \text{for } 1 \leq j \leq K, \\ K & \text{for } K + 1 \leq j \leq L, \\ N - j + 1 & \text{for } L + 1 \leq j \leq N, \end{cases}$$

which is also equal to $w_j^{L,N}$ for L greater than K . \square

Theorem 3.2. *Let \mathbf{F} denote the Hankelized form of the arbitrary $L \times K$ matrix \mathbf{Z} . Then:*

$$T_{\mathbf{Z}-\mathbf{F}}^{L,N} = T_{\mathbf{Z}}^{L,N} - T_{\mathbf{F}}^{L,N},$$

where $T_{\mathbf{Z}}^L = \text{tr}(\mathbf{Z}\mathbf{Z}^T)$.

Proof. It is sufficient to prove that $\text{tr}(\mathbf{Z}\mathbf{F}^T) = \text{tr}(\mathbf{F}\mathbf{F}^T)$.

$$\text{tr}(\mathbf{Z}\mathbf{F}^T) = \sum_{s=1}^{N+1} \sum_{l=s_1}^{s_2} z_{l,s-l} f_{l,s-l} = \sum_{s=1}^{N+1} \sum_{l=s_1}^{s_2} z_{l,s-l} \bar{z}_s = \sum_{s=1}^{N+1} w_{s-1}^{L,N} \bar{z}_s^2 = \text{tr}(\mathbf{F}\mathbf{F}^T),$$

where s_1 and s_2 are defined as above and $w_j^{L,N} = \min\{L, j, N - j + 1\}$. \square

Corollary 3.3. *Let \mathbf{Z} be an arbitrary $L \times K$ matrix and \mathbf{F} be its corresponding Hankelized form. Then: $\text{tr}(\mathbf{Z}\mathbf{Z}^T) \geq \text{tr}(\mathbf{F}\mathbf{F}^T)$.*

Corollary 3.4. *Matrix \mathbf{F} is the nearest matrix to \mathbf{Z} among all the Hankel matrices of dimension $L \times K$ with respect to $T_{\mathbf{Z}}^{L,N}$.*

3.3 Separability

As mentioned earlier, ‘separability’ is the main concept in studying SSA properties. A decomposition of the series Y_N using the SSA method can only be successful if the resulting additive components of the series are approximately separable from each other. The following quantity (called the weighted correlation or *w-correlation*) is a common measure of dependence between the two series $Y_N^{(1)}$ and $Y_N^{(2)}$:

$$\rho_w(Y_N^{(1)}, Y_N^{(2)}) = \frac{\left(Y_N^{(1)}, Y_N^{(2)}\right)_w}{\|Y_N^{(1)}\|_w \|Y_N^{(2)}\|_w}$$

where $\|Y_N^{(1)}\|_w = \sqrt{\left(Y_N^{(1)}, Y_N^{(1)}\right)_w}$, $\|Y_N^{(2)}\|_w = \sqrt{\left(Y_N^{(2)}, Y_N^{(2)}\right)_w}$, $\left(Y_N^{(1)}, Y_N^{(2)}\right)_w = \sum_{j=1}^N w_j y_j^{(1)} y_j^{(2)}$, and $w_j = \min\{L, j, N+1-j\}$ (here we assume $L \leq N/2$) [112].

The matrix of the absolute values of the *w*-correlations is considered as a natural hint for grouping, which corresponds to the full decomposition (each group in this decomposition corresponds to only one matrix component of the SVD). If the absolute value of the *w*-correlations is small, then the corresponding series are almost *w*-orthogonal; however, if its value is large, then those series are far from being *w*-orthogonal and the separability is therefore bad. Thus, if the value of *w*-correlation between two reconstructed components is zero, this means that the two components are separable. If the values of *w*-correlations between reconstructed components are large, this indicates that those components should be gathered into one group and related to the same component in SSA decomposition. In the following, we will provide the results obtained by Hassani et al. [41], which show that the minimum value of the *w*-correlations is obtained at $L = (N+1)/2$, as our empirical results coincide with their results regarding the optimal choice of L .

Theorem 3.5. *For a fixed value of L , let \tilde{S}_L^r be the reconstructed series based on the first r singular values of the trajectory matrix \mathbf{H}_L^r . Then:*

1. $\tilde{S}_L^r = \tilde{S}_K^r$,
2. $\tilde{E}_L^r = \tilde{E}_K^r$, where $\tilde{E}_L^r = Y_N - \tilde{S}_L^r$.

Proof. Here, we only give the proof for the first equality as the second can be easily obtained by the first. Let \tilde{S}_L^r and \tilde{S}_K^r be the constructed series by diagonal averaging of the matrices \mathbf{H}_L^r and \mathbf{H}_K^r , respectively, where:

$$\mathbf{H}_L^r = \sum_{i=1}^r \sqrt{\gamma_i} U_i V_i^T, \quad \mathbf{H}_K^r = \sum_{i=1}^r \sqrt{\gamma_i} V_i U_i^T.$$

Consequently, the results can be attained by equality $\mathbf{H}_L^r = (\mathbf{H}_K^r)^T$. The vectors \tilde{E}_L^r and \tilde{E}_K^r are usually called the noise vectors. The separability between \tilde{S}_L^r and \tilde{E}_L^r (or between \tilde{S}_K^r and \tilde{E}_K^r) is an important task for the reconstruction stage, and also for the forecasting procedure. \square

Corollary 3.6. *Let $\rho_w^{L,r}$ denote the w -correlation between \tilde{S}_L^r and \tilde{E}_L^r . Then, $\rho_w^{L,r} = \rho_w^{K,r}$ which confirms that we need only to consider $L \in \{2, \dots, (N+1)/2\}$.*

Theorem 3.7. *Let \mathbf{H} be the trajectory matrix as defined before and $\mathbf{H} = \mathbf{S} + \mathbf{E} = \tilde{\mathbf{S}} + \tilde{\mathbf{E}}$ where $\mathbf{S} = \sum_{i=1}^r \sqrt{\gamma_i} U_i V_i^T$, $\mathbf{E} = \sum_{j=r+1}^L \sqrt{\gamma_j} U_j V_j^T$, and matrices $\tilde{\mathbf{S}}$ and $\tilde{\mathbf{E}}$ are Hankelized matrices of \mathbf{S} and \mathbf{E} , respectively. Furthermore, consider the following assumptions based on the SVD of the matrices \mathbf{S} and \mathbf{E} for all $i, j = 1, \dots, L$:*

$$U_i U_j^T = \mathbf{0}, \quad V_i V_j^T = \mathbf{0}. \quad (3.7)$$

Then:

$$T_{\tilde{\mathbf{S}}\tilde{\mathbf{E}}^T}^{L,N} = \text{tr}(\tilde{\mathbf{S}}\tilde{\mathbf{E}}^T) > 0. \quad (3.8)$$

Proof. Using the orthogonality feature of eigenvectors with the definition of matrices \mathbf{S} and \mathbf{E} confirm that $\text{tr}(\mathbf{S}\mathbf{E}^T) = 0$. Furthermore, the equality

$$\text{tr}((\mathbf{S} + \mathbf{E})(\mathbf{S} + \mathbf{E})^T) = \text{tr}((\tilde{\mathbf{S}} + \tilde{\mathbf{E}})(\tilde{\mathbf{S}} + \tilde{\mathbf{E}})^T),$$

follows that:

$$tr(\tilde{\mathbf{S}}\tilde{\mathbf{E}}^T) = \frac{(tr(\mathbf{S}\mathbf{S}^T) - tr(\tilde{\mathbf{S}}\tilde{\mathbf{S}}^T)) + (tr(\mathbf{E}\mathbf{E}^T) - tr(\tilde{\mathbf{E}}\tilde{\mathbf{E}}^T))}{2}. \quad (3.9)$$

Let us now show that $tr(\mathbf{S}\mathbf{S}^T) - tr(\tilde{\mathbf{S}}\tilde{\mathbf{S}}^T) = tr(\mathbf{E}\mathbf{E}^T) - tr(\tilde{\mathbf{E}}\tilde{\mathbf{E}}^T)$:

$$\begin{aligned} tr(\mathbf{E}\mathbf{E}^T) - tr(\tilde{\mathbf{E}}\tilde{\mathbf{E}}^T) &= \sum_{s=2}^{N+1} \sum_{l=s_1}^{s_2} (e_{l,s-l} - \bar{e}_s)^2 \\ &= \sum_{s=2}^{N+1} \sum_{l=s_1}^{s_2} \left(e_{l,s-l} - \frac{1}{w_{j-1}^L} \sum_{l=s_1}^{s_2} e_{l,s-l} \right)^2 \\ &= \sum_{j=2}^{N+1} \sum_{i=s_1}^{s_2} \left(y_{j-1} - s_{i,j-i} - \frac{1}{w_{j-1}^L} \sum_{l=s_1}^{s_2} (y_{j-1} - s_{l,j-l}) \right)^2 \\ &= \sum_{j=2}^{N+1} \sum_{i=s_1}^{s_2} (s_{i,j-i} - \bar{s}_j)^2 = tr(\mathbf{S}\mathbf{S}^T) - tr(\tilde{\mathbf{S}}\tilde{\mathbf{S}}^T). \end{aligned}$$

It follows that:

$$tr(\tilde{\mathbf{S}}\tilde{\mathbf{E}}^T) = - \sum_{j=2}^{N+1} \sum_{i=s_1}^{s_2} (s_{i,j-i} - \bar{s}_j) ((e_{i,j-i} - \bar{e}_j)). \quad (3.10)$$

Corollary 3.3 confirms that the right side of Eq. (3.9) is non-negative, which completes the proof. These relations show that if $tr(\tilde{\mathbf{S}}\tilde{\mathbf{S}}^T)$ increases and $tr(\mathbf{S}\mathbf{S}^T) - tr(\tilde{\mathbf{S}}\tilde{\mathbf{S}}^T)$ decreases, then the w -correlation decreases. In addition, the following equality holds:

$$tr(\tilde{\mathbf{S}}\tilde{\mathbf{E}}^T) = tr(\mathbf{E}\mathbf{E}^T) - tr(\tilde{\mathbf{E}}\tilde{\mathbf{E}}^T) = tr(\mathbf{S}\mathbf{S}^T) - tr(\tilde{\mathbf{S}}\tilde{\mathbf{S}}^T).$$

□

Theorem 3.8. $T_{\tilde{\mathbf{S}}_L}^L$ is an increasing function of L on $L \in \{2, \dots, (N+1)/2\}$ provided that there exists a Hankel matrix \mathbf{Z} such that:

$$T_{\mathbf{S}_L - \mathbf{Z}}^L \leq T_{\mathbf{S}_L}^L - T_{\tilde{\mathbf{S}}_{L-m}}^{L-m}, \quad (3.11)$$

where $m \in \{1, \dots, L-2\}$. Thus, the maximum values of these functions are obtained at $L = (N + 1)/2$.

Proof. We will only consider the proof for $T_{\tilde{\mathbf{S}}_L}^L$. Similar proofs and results can be obtained for $T_{\mathbf{E}_L}^L$. Recall from Theorem 3.2 and Corollary 3.4 that for every Hankel matrix \mathbf{Z} of dimension $L \times K$:

$$T_{\mathbf{S}_L}^L - T_{\tilde{\mathbf{S}}_L}^L = T_{\mathbf{S}_L - \tilde{\mathbf{S}}_L}^L \leq T_{\mathbf{S}_L - \mathbf{Z}}^L. \quad (3.12)$$

Let us now assume that \mathbf{Z} is a Hankel matrix that satisfies Eq. (3.11). Consequently, the proof is now completed using Eqs. (3.11) and (3.12). \square

Theorem 3.9. $T_{\mathbf{S}_L - \tilde{\mathbf{S}}_L}^L$ is a decreasing function of L on $L \in \{2, \dots, (N + 1)/2\}$, provided that there exists a Hankel matrix \mathbf{Z} of dimension $L \times K$ such that:

$$T_{\mathbf{S}_L - \mathbf{Z}}^L \leq T_{\mathbf{S}_{L-m}}^{L-m} - T_{\tilde{\mathbf{S}}_{L-m}}^{L-m}, \quad (3.13)$$

where $m \in \{1, \dots, L-2\}$. Thus, the minimum value of this function is obtained at $L = (N + 1)/2$.

Proof. The proof here can be obtained by employing an approach similar to that used in Theorem 3.8. \square

Corollary 3.10. The minimum value of w -correlation attains at $L = (N + 1)/2$, provided that the value of N is large enough and there exists a Hankel matrix \mathbf{Z} of dimension $L \times K$ such that inequality (3.13) is fulfilled.

Proof. To prove this, it is sufficient to show that $T_{\mathbf{S}_L}^{L,N}$ is an increasing function of L . For a large N , we have:

$$T_{\mathbf{S}_L}^{L,N} = \sum_{j=1}^r \gamma_j^{(L,N)} \geq \sum_{j=1}^r \gamma_j^{(L-1,N-1)} \approx T_{\mathbf{S}_{L-1}}^{L-1,N}, \quad (3.14)$$

where $\gamma_j^{L,N}$ denotes the j -th eigenvalue of $\mathbf{H}\mathbf{H}^T$, and \mathbf{H} is the Hankel matrix $L \times K$ corresponding to the series Y_N . Note that every Hankel matrix \mathbf{Z} of dimension $L \times K$ satisfies both inequalities (3.13) and (3.11). Consequently, the proof is completed. \square

It can be noticed from Corollary 3.10 that the reconstructed signal and noise using the leader r eigentriples are almost w -orthogonal for the choice of $L = (N + 1)/2$. A Hankel matrix \mathbf{Z} that satisfies inequality (3.13) can not be easily found, but we can find some equivalent conditions.

Theorem 3.11. *Let $\sigma_l^2(\mathbf{S}_L)$ be the l -th secondary diagonal variance of the matrix \mathbf{S}_L . If $\sigma_l^2(\mathbf{S}_L) \leq \sigma_l^2(\mathbf{S}_{L-m})$, then Theorem 3.9 is satisfied and inequality 3.13 has infinite solutions with respect to \mathbf{Z} .*

Proof. The first part of the theorem is satisfied by using Corollary 3.3 and Theorem 3.2. For the second part, using inequality 3.13, we have:

$$\sum_{j=2}^{N+1} \sum_{i=s_1}^{s_2} (s_{i,j-i}^L - z_{j-1})^2 \leq \sum_{j=2}^{N+1} \sum_{i=s_1}^{s_2} (s_{i,j-i}^{L-m} - \bar{s}_{j-1}^{L-m})^2 = \sum_{j=2}^{N+1} \sigma_{j-1}^2(\mathbf{S}_{L-m}). \quad (3.15)$$

If the following inequality satisfies for $j = 2, \dots, (N + 1)$:

$$\sum_{i=s_1}^{s_2} (s_{i,j-i}^L - z_{j-1})^2 - \sigma_{j-1}^2(\mathbf{S}_{L-m}) \leq 0, \quad (3.16)$$

then; inequality (3.15) is fulfilled. However, the left side of inequality (3.16) is a quadratic form of z_{j-1} , which has the following discriminant:

$$\Delta_{j-1} = \frac{\sigma_{j-1}^2(\mathbf{S}_{L-m}) - \sigma_{j-1}^2(\mathbf{S}_L)}{w_{j-1}^L} \geq 0. \quad (3.17)$$

Thus, it is possible to find infinite real z_{j-1} that satisfies inequality (3.16). This completes the proof. \square

Note that Theorem 3.11 provides enough conditions, as it is obvious from Theorem 3.2 and inequality (3.3) that $T_{\mathbf{S}_L - \tilde{\mathbf{S}}_L}^L \leq T_{\mathbf{S}_{L-m} - \tilde{\mathbf{S}}_{L-m}}^{L-m}$ is equivalent to $\sum_{j=2}^{N+1} \sigma_{j-1}^2(\mathbf{S}_L) \leq \sum_{j=2}^{N+1} \sigma_{j-1}^2(\mathbf{S}_{L-m})$.

3.4 Determinants of a matrix

We consider the determinant of matrix \mathbf{A} because of its importance in the calculation of its eigenvalues. The determinant is a value obtained from a square matrix \mathbf{A} . The determinant of the matrix \mathbf{A} , $\det(\mathbf{A})$ can be calculated from its entries by a particular arithmetic expression, or its value can also be determined in other ways. It is a nonzero number if, and only if, the matrix is invertible. Furthermore, one of its uses is to define the characteristic polynomial of a matrix, which is an essential tool in eigenvalue problems in linear algebra. One of the ways of computing the determinant of a matrix is the Laplace formula.

Before we introduce the Laplace expansion formula, we need to define the minor of a matrix.

Definition 3.12. For any $L \times L$ matrix \mathbf{A} , the determinant $\det(\mathbf{A}_{ij})$ of the $(L-1) \times (L-1)$ matrix obtained from \mathbf{A} by deleting the i -th row and the j -th column is called the (i, j) minor of \mathbf{A} . Using minors we can show one way to calculate the determinant of an $L \times L$ matrix. The technique is called Laplace expansion by cofactors.

Definition 3.13 (The Laplace Expansion formula). The determinant of an $L \times L$ matrix $\mathbf{A} = (a_{ij})$, where $L \geq 2$, is a scalar and can be computed as

$$\det(\mathbf{A}) = a_{i1}c_{i1} + a_{i2}c_{i2} + \cdots + a_{iL}c_{iL} = \sum_{j=1}^L a_{ij}c_{ij}, \quad i = 1, \dots, L \quad (3.18)$$

which is the cofactor expansion along the i -th row, and also as

$$\det(\mathbf{A}) = a_{1j}c_{1j} + a_{2j}c_{2j} + \cdots + a_{Lj}c_{Lj} = \sum_{i=1}^L a_{ij}c_{ij}, \quad (3.19)$$

the cofactor expansion along the j -th column. Since

$$c_{ij} = (-1)^{i+j} \det(\mathbf{A}_{ij}), \quad (3.20)$$

each cofactor is plus or minus the corresponding minor, with the correct sign given by the term $(-1)^{i+j}$. Note that this procedure reduces the problem of calculating the determinant of an $L \times L$ matrix to the problem of computing the determinant of an $(L-1) \times (L-1)$ matrix. Using this technique can minimise the main problem to one of computing the determinants of 2×2 matrices.

3.5 Theoretical results

3.5.1 The problem

Consider a one-dimensional series $Y_N = (y_1, \dots, y_N)$ of length N . Let $Y_N = S_N + E_N$, where S_N represents the signal component and E_N represents the noise component. Transferring this series into the multi-dimensional series H_1, \dots, H_K with vectors $H_i = (y_i, \dots, y_{i+L-1})^T \in \mathbb{R}^L$ provides the following trajectory Hankel matrix $\mathbf{H} = (h_{i,j})_{i,j=1}^{L,K}$. It is clear that $\mathbf{H} = \mathbf{S} + \mathbf{E}$, where \mathbf{S} and \mathbf{E} respectively represent Hankel matrices of the signal S_N and noise E_N series. Note that L is in fact an embedding dimension, the matrix $\mathbf{H}\mathbf{H}^T$ is similar to a correlation matrix, and the procedure of constructing the Hankel matrix is a special case of delay embedding. There exists a huge volume of research devoted to the analysis of experimental data by means of the tools of nonlinear dynamics, and in relation to the crucial concept of embedding, including the delayed embedding which we use in this thesis (see for example [184, 185]).

The Hankel matrix \mathbf{H} and its corresponding singular values play a pivotal role in various fields including time series analysis [107, 186], biomedical signal processing [27, 43, 118], mathematics [37], econometrics [187] and physics [188]. The distribution of eigenvalues/singular values and their closed form are of great interest,

but this issue has not been considered adequately [189]. We have already mentioned that the singular values of the trajectory matrix \mathbf{H} are the square root of the eigenvalues of the L by L matrix $\mathbf{A} = \mathbf{H}\mathbf{H}^T$, where $H_i = U_i\sqrt{\gamma_i}V_i^T$, U_i is the eigenvector corresponding to the eigenvalue γ_i ($i = 1, \dots, L$) of \mathbf{A} ($L \leq K$), and $V_i = \mathbf{H}^T U_i / \sqrt{\gamma_i}$. The ratio of each eigenvalue γ_i is the contribution of the matrix \mathbf{H}_i to \mathbf{H} , since $\|\mathbf{H}\|_F^2 = \sum_{i=1}^L \gamma_i$ and $\|\mathbf{H}_i\|_F^2 = \gamma_i$. We also mentioned that for a fixed value of L ; $L \leq K$, and a series with length N , the trace of \mathbf{A} , $tr(\mathbf{A})$, is $T_{\mathbf{H}}^{L,N} = \|\mathbf{H}\|_F^2 = tr(\mathbf{A}) = \sum_{i=1}^L \gamma_i$.

Let us now assume that \mathbf{X} is an $L \times v$ rectangular matrix whose entries are independent identically distributed random variables, then the matrix \mathbf{X} will be called a random matrix. Currently, the random matrix theory (RMT) (or the spectral analysis of large dimensional random matrices), is the only methodical theory that can be used as a powerful tool to solve problems in the analysis of large dimensional data, which has become a hot topic in statistics and diverse disciplines in which statistics is applicable. The roots of RMT go back to the development of quantum mechanics (QM) in the 1940's and 1950's. In 1955, [190] proved and established the well known semi-circular law. Since then, researchers from mathematics, physics and statistics fields have developed the concepts and results of random matrix theory. The statistics show that the number of publications on random matrix theory has been rapidly increasing [191].

A brief review of Random Matrix Theory

As mentioned above, RMT can be traced back to the the 1940's and early 1950's in the development of quantum mechanics [191]. The eigenvalues of a Hermitian operator on a Hilbert space, which is called Hamiltonian, describe the energy levels of a system in QM. For more information on RMT and its applications in various areas, we refer to [192]–[194].

In [190, 195], it was proved that the anticipated spectral distribution of a large dimensional Winger matrix tends to semicircular law. Subsequently, this research was developed and generalised in different aspects (see [196, 197]). It was also proved that if the sample size is greater than the dimension of the Winger matrix,

then the spectral distribution of a sample covariance matrix tends to semicircular law [198]. Furthermore, Marchenko and Pastur [199], and Pastur [200, 201] studied the asymptotic theory of spectral analysis of large dimensional sample covariance matrices; this was developed by many authors (see [202–206]). In addition, the limiting spectral distribution of products of random matrices were investigated in [202, 206–209]. Recently, studies on RMT have focused on limiting theorems; for example, the central limit theorem and limiting distributions of spectral spacings and largest eigenvalues [191].

Suppose now that $\mathbf{P} = \mathbf{X}\mathbf{X}^T$ is an $L \times L$ matrix with eigenvalues λ_i ($i = 1, \dots, L$), and all the eigenvalues are real, then the one-dimensional distribution function can be defined as follows:

$$F^{\mathbf{P}}(y) = \frac{1}{L} \# \{i \leq L : \lambda_i \leq y\}.$$

This function is called the empirical spectral distribution (ESD) of the matrix \mathbf{X} , and $\#$ denotes the cardinality of the set E .

For a given sequence of random matrices \mathbf{P}_L , random matrix theorem is used to study the convergence of the sequence of empirical spectral distributions $F^{\mathbf{P}_L}$. The limit distribution F is called the limiting spectral distribution (LSD) of those random matrices. It is also worth mentioning that ESD is important in multivariate statistics because many important statistics can be expressed as functions of the ESD of some random matrices.

If \mathbf{P} is an $L \times L$ positive definite matrix, then we can define the determinant of \mathbf{P} as follows.

$$\det(\mathbf{P}) = \prod_{i=1}^L \lambda_i = \exp\left(L \int_0^\infty \log y F^{\mathbf{P}}(dy)\right). \quad (3.21)$$

Let us return back to the rectangular random matrix \mathbf{X} , and X_i , a random vector in \mathbf{X} . It is defined that X_i is a random vector from an L -variate normal distribution with zero mean if every linear combination of its L components has a univariate

normal distribution, which can be written in the following notation [193], [210, 211];

$$X_i = (y_i, \dots, y_{i+L-1})^T \sim N_L(0, \mathbf{G}), \quad (3.22)$$

where 0 is an L -dimensional mean vector, and \mathbf{G} is an $L \times L$ covariance matrix. The multivariate normal distribution is often used to describe any set of possibly correlated real-valued random variables, each of which clusters around a mean value. Note that every covariance matrix is positive semi-definite. It is also worth mentioning that the importance of a random vector is derived from the multivariate central limit theorem.

The sample covariance forms often lead to the Wishart distribution [212], which is a generalisation of the univariate chi-square distribution to multivariables. It is of great importance in the estimation of covariance matrices; it is parameterised for these matrices, the diagonal elements of which are each chi-square random variables. This type of distribution is often used as a model for the distribution of the sample covariance matrix for multivariate random normal data, after scaling the data by its size.

Then, the Wishart distribution is the probability distribution of the $L \times L$ random matrix $\mathbf{X}\mathbf{X}^T$:

$$\mathbf{X}\mathbf{X}^T \sim W_L(\mathbf{G}, v), \quad (3.23)$$

where the positive integer v is the number of degrees of freedom that are analogous to the degrees of freedom parameter of a univariate chi-square distribution (see, for example, [213]). For $L \leq v$ the matrix $\mathbf{X}\mathbf{X}^T$ is invertible with probability 1 if \mathbf{G} is invertible. Note that when $L = 1$, this distribution is a chi-squared distribution with v degrees of freedom.

Based on the limit central theorem, asymptotic results in the following theorem for large samples provide beneficial distributions for the eigenvalues and eigenvectors of the sample covariance matrix.

Theorem 3.14. *Let \mathbf{G} be a positive-definite matrix with distinct eigenvalues. Let $\mathbf{X}\mathbf{X}^T \sim W_L(\mathbf{G}, v)$ where $\mathbf{X}(L \times v)$, and set $\mathbf{J} = v^{-1}\mathbf{X}\mathbf{X}^T$. Consider spectral decomposition $\mathbf{G} = \mathbf{Z}\mathbf{\Lambda}\mathbf{Z}^T$ and $\mathbf{J} = \mathbf{Q}\mathbf{\Gamma}\mathbf{Q}^T$, and let η and φ be the vectors of diagonal elements in $\mathbf{\Lambda}$ and $\mathbf{\Gamma}$. Consequently, the following asymptotic distributions hold as $v \rightarrow \infty$:*

- $\varphi \sim N_L(\eta, 2\mathbf{\Lambda}^2/v)$, where the eigenvalues of \mathbf{J} are asymptotically normal, unbiased, and independent, with φ_i having variance $2\eta_i^2/v$.
- $Q_i \sim N_L(\Upsilon_i, \mathbf{C}/v)$, where

$$C_i = \varphi_i \sum_{i \neq j} \frac{\varphi_i}{(\varphi_i - \varphi_j)^2} \Upsilon_i \Upsilon_i^T, \quad (3.24)$$

that is, the eigenvectors of \mathbf{J} are asymptotically normal, unbiased, and have the stated asymptotic covariance matrix \mathbf{C}/v .

- The covariance between the r -th element of Q_i and the s -th element of Q_j is $-\varphi_i \varphi_j \Upsilon_{rj} \Upsilon_{si} / v (\varphi_i - \varphi_j)^2$.
- The elements of φ are asymptotically independent of the elements of \mathbf{Q} [214].

We will now turn our attention to Marchenko-Pastur distribution [199]. Marchenko-Pastur distribution describes the asymptotic behavior of singular values of large rectangular random matrices. Let \mathbf{X} and \mathbf{J} be as above, where x_{ij} are independent and identically distributed (iid), $E(x_{ij}) = 0$ and $E(x_{ij}^2) = 1$, and let $\varphi_1, \dots, \varphi_L$ be the eigenvalues of \mathbf{J} . Then, consider the random spectral measure:

$$\varpi_v = \frac{1}{L} \sum_{i=1}^L \delta_{\varphi_i}. \quad (3.25)$$

We can now state the Marchenko-Pastur distribution as in [191].

Theorem 3.15. *Assume that \mathbf{J} follows $W(\mathbf{I}/v, v)$, where \mathbf{I} is the $L \times L$ identity matrix. If L and v go to infinity in such a way that L/v leads $y \in (0, 1)$ then $\mu_v \rightarrow$*

ϖ where ϖ is a deterministic measure which converges weakly to the Marchenko-Pastur distribution with density:

$$\frac{d\mu}{dx} = \frac{\sqrt{(a_2 - x)(x - a_1)}}{2\pi xy} 1_{(a_1 \leq x \leq a_2)}, \quad (3.26)$$

where

$$a_1 = (1 - \sqrt{y})^2, \quad a_2 = (1 + \sqrt{y})^2. \quad (3.27)$$

The above theorems hold for the situation in which the vectors X_i are distributed independently, whilst for the Hankel matrix this is not applicable, as the lagged vectors H_i and H_j are correlated. For example, H_i and H_{i+1} ($i = 1, \dots, K - 1$) have $L - 1$ similar observations with the following covariance matrix:

$$\text{Cov}(H_i, H_{i+1}) = \sigma^2 \begin{pmatrix} \mathbf{0}_{1 \times (L-1)} & \mathbf{0}_{(L-1) \times 1} \\ \mathbf{I}_{(L-1) \times (L-1)} & 0 \end{pmatrix} \quad (3.28)$$

where σ^2 is the variance of y_i , \mathbf{I} is the identity matrix or unit matrix of size $L - 1$ with ones on the main diagonal and zeros elsewhere, and $\mathbf{0}$ is a vector of size $L - 1$ with all its entries zero.

Moreover, it is always of interest to have bounded eigenvalues, whilst in the above case the magnitude of eigenvalues changes with the series length; increasing the sample size N leads to an increase in γ_i .

3.5.2 The solution

To overcome the issue above, we divide \mathbf{A} by its trace $\left(\sum_{i=1}^L \gamma_i\right)$. This in turn provides several important properties.

Proposition 1. Let ζ_1, \dots, ζ_L denote the eigenvalues of the matrix $\mathbf{B} = \mathbf{A} / \sum_{i=1}^L \gamma_i$, where $\mathbf{A} = \mathbf{H}\mathbf{H}^T$, and γ_i ($i = 1, \dots, L$) are the eigenvalues of \mathbf{A} . In this case, we have the following properties:

1. $0 \leq \zeta_L \leq \dots \leq \zeta_1 \leq 1$,
2. $\sum_{i=1}^L \zeta_i = 1$,
3. $\zeta_1 \geq \frac{1}{L}$,
4. $\zeta_L \leq \frac{1}{L}$,
5. $\zeta_i \in (\frac{1}{L} - a, \frac{1}{L} + b)$ ($i = 2, \dots, L-1$), where $a, b \in [0, 1]$.

Proof. The first property is simply obtained from matrix algebra and the fact that there is a rational number between two real numbers. \square

Theorem 3.16. $(\forall \zeta_1), (\forall \zeta_L), (\zeta_1 > 0 \text{ and } \zeta_L > 0) \Rightarrow (\exists \zeta_i) (\zeta_i \text{ is rational and } (\zeta_L < \zeta_i < \zeta_1))$.

Another way to prove the first property:

Proof. $\because \gamma_i \geq 0$ and $\zeta_i = \frac{\gamma_i}{\sum \gamma_i}$,

$\therefore \zeta_i \geq 0$.

Furthermore,

$$\gamma_1 \geq \gamma_2 \geq \dots \geq \gamma_L \geq 0 \Rightarrow \frac{\gamma_1}{\gamma_1} \geq \frac{\gamma_i}{\gamma_1} \Rightarrow 1 \geq \frac{\gamma_i}{\sum \gamma_i} = \zeta_i.$$

Thus, $0 \leq \zeta_i \leq 1$.

To prove the second property, We have

$$\zeta_1 = \frac{\gamma_1}{\sum \gamma_i}, \zeta_2 = \frac{\gamma_2}{\sum \gamma_i}, \dots, \zeta_L = \frac{\gamma_L}{\sum \gamma_i} \Rightarrow \zeta_1 + \zeta_2 + \dots + \zeta_L = \frac{\sum \gamma_i}{\sum \gamma_i} = 1.$$

To prove the third property, the first two properties are used as follows. The second part confirms

$$\zeta_1 + \zeta_2 + \dots + \zeta_L = 1.$$

Thus, using the first property, $\zeta_1 \geq \zeta_i$ ($i = 2, \dots, L$), we obtain

$$\underbrace{\zeta_1 + \zeta_1 + \dots + \zeta_1}_{L \text{ elements}} = L\zeta_1 \geq 1 \Rightarrow \zeta_1 \geq 1/L.$$

Similarly, for the fourth property, it is straightforward to show that

$$\underbrace{\zeta_L + \zeta_L + \dots + \zeta_L}_{L \text{ elements}} = L\zeta_L \leq 1 \Rightarrow \zeta_L \leq 1/L,$$

since $\zeta_L \leq \zeta_i$, $i = (1, 2, \dots, L-1)$, and $\sum_{i=1}^L \zeta_i = 1$.

To prove part 5, let us first prove that there exists ζ_2 between real numbers ζ_1 and ζ_L . It is clear that $\zeta_L < \zeta_1$ for $L > 2$. Since $\zeta_1 - \zeta_L > 0$, we can then choose a natural number n , large enough to make $\frac{1}{n} < \zeta_1 - \zeta_L$. Now, from the numbers $\frac{1}{n}, \frac{2}{n}, \dots, \frac{k}{n}$ we select the largest possible natural number k such that $\frac{k}{n} \leq \zeta_L$. Therefore, $\zeta_L < \frac{k+1}{n}$. Note that $\frac{k+1}{n} < \zeta_1$ since if we assume $\frac{k+1}{n} \leq \zeta_1$ then $\frac{1}{n} = \frac{k+1}{n} - \frac{k}{n} \geq \zeta_1 - \zeta_L$, which is false as n was picked such that $\frac{1}{n} < \zeta_1 - \zeta_L$. Thus, $\zeta_2 = \frac{k+1}{n}$ satisfies $\zeta_L < \zeta_2 < \zeta_1$. This approach can be used for other ζ_i . \square

Let us first compare the distribution of the eigenvalues, λ_i , of the matrix $\mathbf{X}\mathbf{X}^T$, and the distribution of the eigenvalues, ζ_i , of \mathbf{B} for white noise series of length $N = 10^5$. For the purpose of visualisation, let us consider $L = 10$ for $m = 5 \times 10^3$ simulation. Note that the dimension of \mathbf{X} is 10×10^4 , whilst 10×99991 for \mathbf{H} as previously defined. Figure. 3.1 (left) illustrates the distribution of λ_i ($i = 1, 5, 10$) for $\mathbf{X}\mathbf{X}^T$, and Figure. 3.1 (right) shows the distribution of ζ_i ($i = 1, 5, 10$) for \mathbf{B} . It can be seen that there are some similarities and dissimilarities. For example, the distributions of λ_1 and ζ_1 are skewed to the right, whilst the distribution tail for λ_1 is shorter than the case of ζ_1 ; the skewness coefficient for ζ_1 is greater than the corresponding value for λ_1 . Furthermore, for the middle ζ_i and λ_i , particularly $i = 5$, both distributions are symmetric.

The results indicate that the distribution of the first and last eigenvalues of \mathbf{B} tend to have a skewed distribution, whilst the middle eigenvalue may have an asymptotically symmetric distribution. Let us now evaluate the asymptotical behaviour of the largest and smallest eigenvalues, for different values of N , generated from a white

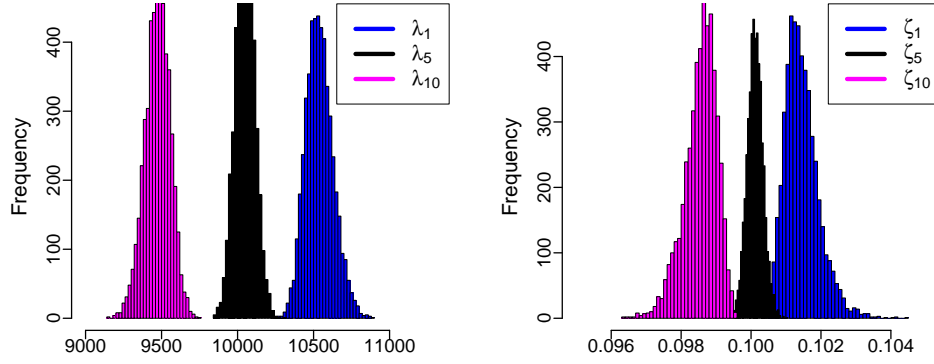


FIGURE 3.1: Histogram of ζ_i ($i = 1, 5, 10$) for matrix \mathbf{B} (right), and the histogram of λ_i ($i = 1, 5, 10$) for matrix \mathbf{XX}^T (left).

noise series. Figure. 3.2 (left) displays the results for ζ_i ($i = 1, 10$), and Figure. 3.2 (right) shows the the results for λ_i ($i = 1, 10$).

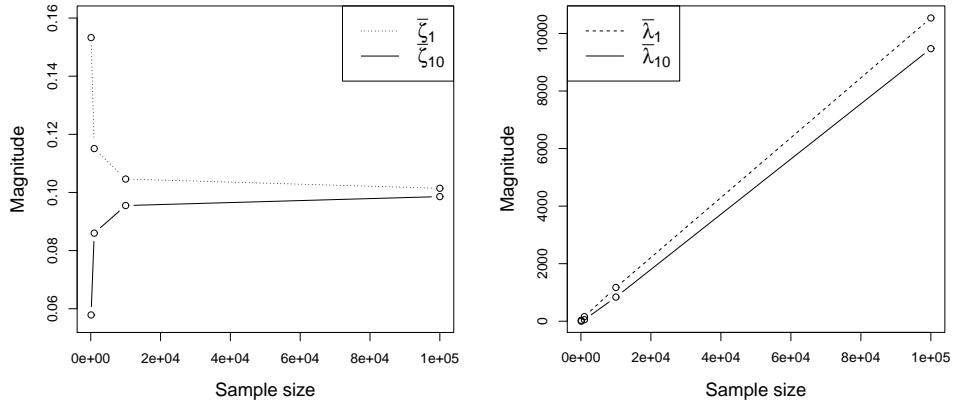


FIGURE 3.2: Plot of $\bar{\zeta}_i$ and $\bar{\lambda}_i$ ($i = 1, 10$) for different sample size N for a white noise series.

As the results indicate, the gap between ζ_1 and ζ_{10} becomes smaller as the sample size N increases and both asymptotically converge to $\frac{1}{L}$. Thus, according to the above proposition, other ζ_i tend to $\frac{1}{L}$. However, both λ_1 and λ_{10} increase following an increase in sample size N and the discrepancy between them becomes larger. In this research, we only consider ζ_i as there is no adequate information about the empirical distribution of ζ_i , whilst there is much about λ_i (see for example [193, 199, 215]).

Proposition 2. For the \mathbf{A} and \mathbf{B} symmetric matrices, where $\mathbf{A} = \mathbf{H}\mathbf{H}^T$ and $\mathbf{B} = \mathbf{A}/\text{tr}(\mathbf{A})$, then we have:

i) $\log(\det(\exp(\mathbf{B}))) = 1$

ii) $\det(\mathbf{B}) < \frac{\text{tr}(\mathbf{B})}{L} = \frac{1}{L}$

Proof. i) We have ζ_1, \dots, ζ_L , the eigenvalues of \mathbf{B} . Let $\mathbf{D} = \exp(\mathbf{B})$, then:

$$\mathbf{D} = \exp(\mathbf{B}) = \mathbf{U} \text{diag}(e^{\zeta_1}, \dots, e^{\zeta_L}) \mathbf{U}^{-1},$$

where $d_1 = e^{\zeta_1}, \dots, d_L = e^{\zeta_L}$ are the eigenvalues of \mathbf{D} .

Consequently, the determinant of \mathbf{D} is the product of all the eigenvalues, which can be written as follows,

$$\det(\mathbf{D}) = \prod_{i=1}^L d_i = \prod_{i=1}^L e^{\zeta_i} = e^{\zeta_1} e^{\zeta_2} \dots e^{\zeta_L},$$

where $\prod_{i=1}^L$ is notation for the product. Now take the logarithm

$$\log(\det(\mathbf{D})) = \log\left(\prod_{i=1}^L d_i\right) = \log\left(\prod_{i=1}^L e^{\zeta_i}\right) = \log(e^{\zeta_1} e^{\zeta_2} \dots e^{\zeta_L})$$

This yields

$$\begin{aligned} \log(\det(\mathbf{D})) &= \log(e^{\zeta_1}) + \log(e^{\zeta_2}) + \dots + \log(e^{\zeta_L}) \\ &= \zeta_1 + \zeta_2 + \dots + \zeta_L = \sum_{i=1}^L \zeta_i = \text{tr}(\mathbf{B}) = 1 \end{aligned}$$

Another way to prove $1 = \text{tr}(\mathbf{B}) = \log(\det(\exp(\mathbf{B})))$:

From Proposition 1, we have:

$$\sum_{i=1}^L \zeta_i = 1,$$

$$\Rightarrow e^{\sum_{i=1}^L \zeta_i} = e^{\zeta_1 + \zeta_2 + \dots + \zeta_L} = e^{\zeta_1} e^{\zeta_2} \dots e^{\zeta_L} = \prod_{i=1}^L e^{\zeta_i} = \det(\exp(\mathbf{B})),$$

$$\Rightarrow e^{\sum_{i=1}^L \zeta_i} = \det(\exp(\mathbf{B})).$$

Take the logarithm for both sides, that is

$$\log \left(e^{\sum_{i=1}^L \zeta_i} \right) = \log (\det(\exp((B)))) ,$$

$$\Rightarrow \sum_{i=1}^L \zeta_i = \log (\det(\exp((B)))) .$$

We know that $\text{tr}(\mathbf{B}) = \sum_{i=1}^L \zeta_i$. This yields:

$$\text{tr}(\mathbf{B}) = \log (\det(\exp(\mathbf{B}))) = 1 .$$

ii) From Proposition 1, we have $\zeta_L \leq \frac{1}{L}$ and $0 \leq \zeta_L \dots \zeta_1 \leq 1$, thus

$$\prod_{i=1}^L \zeta_i < \zeta_L \Rightarrow \prod_{i=1}^L \zeta_i < \frac{1}{L} \Rightarrow \det(\mathbf{B}) < \frac{1}{L} .$$

We also have

$$\text{tr}(\mathbf{B}) = \sum_{i=1}^L \zeta_i = 1 . \text{ This yields}$$

$$\det(\mathbf{B}) < \frac{\text{tr}(\mathbf{B})}{L} .$$

□

It is known that there is a close connection between determinants, eigenvalues, rank, and invertibility for a square matrix. Note that the determinant of matrix \mathbf{B} is the product of its eigenvalues. It should also be mentioned that an $L \times L$ matrix \mathbf{B} has an inverse if and only if $\text{rank}(\mathbf{B}) = L$. Based on the following theorem we know that the determinant of \mathbf{B} is zero if and only if when its row reduced the resulting matrix has a row of zeros; this happens when its rank is less than L .

Theorem 3.17. *The determinant of an $L \times L$ matrix \mathbf{B} is nonzero if and only if its rank is L ; that is to say, if and only if it is invertible [216].*

Corollary 3.18. *If Y_N is a white noise process and N is large, we have $\zeta_i \approx 1/L$. Consequently,*

$$\det(\mathbf{B}) = \prod_{i=1}^L \zeta_i \approx \frac{1}{L^L} . \tag{3.29}$$

Corollary 3.19. *Based on Theorem 3.17, the value of the determinant of \mathbf{B} for the white noise process is greater than its value for any other case. Thus, the general form of the determinant of matrix \mathbf{B} can be defined as:*

$$\det(\mathbf{B}) = \prod_{i=1}^L \zeta_i < \frac{1}{L^L}. \quad (3.30)$$

The proposed solution and results obtained above indicate that the distribution of the eigenvalues and their related form such as skewness and kurtosis can help to distinguish between signal and noise components, then identifying the eigenvalues corresponding to the signal subspace.

3.5.3 The general form of ζ_i

Consider the trajectory matrix \mathbf{H} defined in Eq. (1.1), with window length L . In this case $\mathbf{B} = \mathbf{A}/\text{tr}(\mathbf{A})$, where $\mathbf{A} = \mathbf{H}\mathbf{H}^T$ is an $L \times L$ symmetric matrix with the following structure:

$$\mathbf{B} = \frac{1}{\text{tr}(\mathbf{A})} \begin{pmatrix} \sum_{i=1}^K y_i^2 & \sum_{j=2}^{K+1} \sum_{i=1}^K y_i y_j & \cdots & \sum_{j=L}^N \sum_{i=1}^K y_i y_j \\ \sum_{j=2}^{K+1} \sum_{i=1}^K y_i y_j & \sum_{i=2}^{K+1} y_i^2 & \cdots & \sum_{j=L}^N \sum_{i=L}^{K+1} y_i y_j \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{j=L}^N \sum_{i=1}^K y_i y_j & \sum_{j=L}^N \sum_{i=2}^{K+1} y_i y_j & \cdots & \sum_{i=L}^N y_i^2 \end{pmatrix}. \quad (3.31)$$

To find the the eigenvalue of the matrix \mathbf{B} , one needs to find the following determinant:

$$f(\zeta) = \det(\mathbf{B} - \zeta \mathbf{I}) \quad (3.32)$$

$$= \det \begin{pmatrix} \sum_{i=1}^K y_i^2 / \text{tr}(\mathbf{A}) - \zeta & \sum_{j=2}^{K+1} \sum_{i=1}^K y_i y_j / \text{tr}(\mathbf{A}) & \cdots & \sum_{j=L}^N \sum_{i=1}^K y_i y_j / \text{tr}(\mathbf{A}) \\ \sum_{j=2}^{K+1} \sum_{i=1}^K y_i y_j / \text{tr}(\mathbf{A}) & \sum_{i=2}^{K+1} y_i^2 / \text{tr}(\mathbf{A}) - \zeta & \cdots & \sum_{j=L}^N \sum_{i=L}^{K+1} y_i y_j / \text{tr}(\mathbf{A}) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{j=L}^N \sum_{i=1}^K y_i y_j / \text{tr}(\mathbf{A}) & \sum_{j=L}^N \sum_{i=2}^{K+1} y_i y_j / \text{tr}(\mathbf{A}) & \cdots & \sum_{i=L}^N y_i^2 / \text{tr}(\mathbf{A}) - \zeta \end{pmatrix} \quad (3.33)$$

Therefore, the characteristic polynomial is:

$$f(\zeta) = (-1)^L [\zeta^L + c_1 \zeta^{L-1} + c_2 \zeta^{L-2} + \dots + c_{L-1} \zeta + c_L], \quad (3.34)$$

where c_i are the coefficients to be computed by evaluating the determinant using the cofactor expansion explained earlier. Note that Eq. (3.34) is correct for any value of ζ . If we set $\zeta = 0$, then Eq. (3.34) yields:

$$f(0) = \det(\mathbf{B}) = (-1)^L c_L.$$

Thus,

$$c_L = (-1)^L \det(\mathbf{B}).$$

Note also that c_1 can be obtained by evaluating the determinant in Eq. (3.34) using also the cofactor expansion, which is the trace of \mathbf{B} .

$$c_1 = -\text{tr}(\mathbf{B}) = -1.$$

Therefore, the general form for the characteristic polynomial is:

$$f(\zeta) = (-1)^L [\zeta^L - \zeta^{L-1} + c_2 \zeta^{L-2} + \dots + (-1)^{L-1} c_{L-1} \zeta + (-1)^L \det(\mathbf{B})], \quad (3.35)$$

where $\det(\mathbf{B}) = \det(\mathbf{A}/\text{tr}(\mathbf{A})) = \det(\mathbf{A})/\text{tr}^2(\mathbf{A})$.

Theorem 3.20. *Let \mathbf{A} and \mathbf{B} are defined as above, then the eigenvalues ζ_i of \mathbf{B} are all real.*

Proof. Suppose $\zeta \in \mathbb{C}$ is an eigenvalue of \mathbf{B} and let $U \in \mathbb{C}^L$ be its corresponding eigenvector, then we can write:

$$\mathbf{B}U = \zeta U, \quad (3.36)$$

where \mathbb{C} is the set of the complex numbers $z = x + iy$ where x and y are the real and imaginary parts of z and $i = \sqrt{-1}$. Consequently, \mathbb{C}^L is the set of L -column

vectors with components in \mathbb{C} , and similarly $\mathbb{C}^{L \times L}$ is the set of $L \times L$ matrices with complex numbers as its entries. The complex conjugate of z is written as $z^* = x - iy$. Similarly for $U \in \mathbb{C}^L$ and $\mathbf{B} \in \mathbb{C}^{L \times L}$, their complex conjugates are $U^* \in \mathbb{C}^L$ and $\mathbf{B}^* \in \mathbb{C}^{L \times L}$, which are obtained by considering the complex conjugate of each of their components.

Now take the complex conjugates of both sides of Eq. (3.36); this gives:

$$\mathbf{B}^* U^* = \zeta^* U^*. \quad (3.37)$$

Now we can pre-multiply (3.36) with $(U^*)^T$:

$$\begin{aligned} \zeta (U^*)^T U &= (U^*)^T (\mathbf{B} U) = ((U^*)^T \mathbf{B}) U \\ &= (\mathbf{B}^T U^*)^T U \quad \text{since } (\mathbf{B} U)^T = U^T \mathbf{B}^T \\ &= (\mathbf{B} U^*)^T U \quad \text{since } \mathbf{B}^T = \mathbf{B} \\ &= (\zeta^* U^*)^T U = \zeta^* (U^*)^T U_i \quad \text{using (Eq. 3.37)}. \end{aligned}$$

Thus,

$$(\zeta - \zeta^*) (U^*)^T U = 0.$$

However, U , being an eigenvector is non-zero and $(U^*)^T U = \sum_{i=1}^L u_i^* u_i > 0$ since at least one of the components of U is non-zero and for any complex number $z = x + iy$, $z^* z = x^2 + y^2 \geq 0$. Hence, $\zeta = \zeta^*$, i.e., ζ and hence U are real. \square

Theorem 3.21. *Let U_1 and U_2 be eigenvectors for the real symmetric matrix $\mathbf{B} = \mathbf{A}/\text{tr}(\mathbf{A})$ corresponding to the different eigenvalues, ζ_1 and ζ_2 . Consequently, U_1 and U_2 are orthogonal.*

Proof. The eigenvectors U_1 and U_2 of \mathbf{B} satisfy the following equations:

$$\mathbf{B} U_1 = \zeta_1 U_1, \quad \mathbf{B} U_2 = \zeta_2 U_2 \quad \text{and } \zeta_1 \neq \zeta_2.$$

Therefore,

$$\zeta_1 U_2^T U_1 = U_2^T (\mathbf{B} U_1) = (U_2^T \mathbf{B}) U_1 = (\mathbf{B}^T U_2)^T U_1 = (\mathbf{B} U_2)^T U_1 = \zeta_2 U_2^T U_1.$$

$\Rightarrow (\zeta_1 - \zeta_2) U_2^T U_1 = 0$. Since $\zeta_1 - \zeta_2 \neq 0$, then $U_2^T U_1 = 0$. Thus, they are orthogonal. \square

3.5.4 The eigenvalues and determinant of \mathbf{B} for $L = 2$

Let us now consider the situation where $L=2$. In this case \mathbf{B} is a square-symmetric matrix with the following structure:

$$\mathbf{B} = \frac{1}{y_1^2 + y_N^2} + 2 \sum_{i=2}^{N-1} y_i^2 \begin{pmatrix} \sum_{i=1}^{N-1} y_i^2 & \sum_{i=1}^{N-1} \sum_{j=2}^N y_i y_j \\ \sum_{i=1}^{N-1} \sum_{j=2}^N y_i y_j & \sum_{i=2}^N y_i^2 \end{pmatrix}. \quad (3.38)$$

To find the eigenvalue of matrix \mathbf{B} , we use Eq. (3.35) with $L = 2$. This yields:

$$f(\zeta) = \zeta^2 - \zeta + \det(\mathbf{B}). \quad (3.39)$$

Thus, the eigenvalues can be obtained using the quadratic formula:

$$\zeta = \frac{1 \pm \sqrt{1 - 4 \det(\mathbf{B})}}{2}.$$

Consequently, the eigenvalues of \mathbf{B} , ζ_1 and ζ_2 , are as follows:

$$\zeta_1 = \frac{1 + \sqrt{1 - 4 \det(\mathbf{B})}}{2} = \frac{1}{2} + \frac{1}{2} \sqrt{1 - \frac{4 \det(\mathbf{A})}{\text{tr}^2(\mathbf{A})}}, \quad (3.40)$$

$$\zeta_2 = \frac{1 - \sqrt{1 - 4 \det(\mathbf{B})}}{2} = \frac{1}{2} - \frac{1}{2} \sqrt{1 - \frac{4 \det(\mathbf{A})}{\text{tr}^2(\mathbf{A})}}. \quad (3.41)$$

Corollary 3.22. *Using the first property of Proposition 1, we obtain:*

$$\zeta_1 = \frac{1}{2} + \frac{1}{2} \sqrt{1 - \frac{4 \det(\mathbf{A})}{\text{tr}^2(\mathbf{A})}} < 1.$$

Therefore,

$$4 \det(\mathbf{B}) = 4 \det(\mathbf{A}) / \text{tr}^2(\mathbf{A}) > 0.$$

We also have:

$$1 - \frac{4 \det(\mathbf{A})}{\text{tr}^2(\mathbf{A})} > 0,$$

$$\Rightarrow I_{\mathbf{A}} = 4 \det(\mathbf{B}) = 4 \det(\mathbf{A}) / \text{tr}^2(\mathbf{A}) < 1,$$

$$\Rightarrow I_{\mathbf{A}} = 4 \det(\mathbf{B}) \in (0, 1). \quad (3.42)$$

Corollary 3.23. *Note that if Y_N is a white noise process, then matrix \mathbf{H} is fully ranked. Thus, for a large N :*

$$\det(\mathbf{B}) = \frac{\det(\mathbf{A})}{\text{tr}^2(\mathbf{A})} \rightarrow \frac{1}{4}.$$

$\Rightarrow \zeta_1 \gtrapprox 1/2$, and $\zeta_2 \lesssim 1/2$. But, if Y_N is a trend, for example where \mathbf{H} has rank one, then

$$\det(\mathbf{B}) = \frac{\det(\mathbf{A})}{\text{tr}^2(\mathbf{A})} \rightarrow 0 \Rightarrow \zeta_1 \lesssim 1, \zeta_2 \gtrapprox 0.$$

Corollary 3.24. *For a white noise process:*

$$4 \det(\mathbf{B}) = \frac{4 \det(\mathbf{A})}{\text{tr}^2(\mathbf{A})} \approx \log(\det(\exp(\mathbf{B}))) = 1$$

The above results confirm that for $L = 2$, ζ_1 has a positive skewed distribution whilst ζ_2 has a negative skewed one for a white noise process. This helps us to distinguish between signal and noise and to determine the number of eigenvalues required for signal extraction.

The eigenvalues of \mathbf{D}

To find the eigenvalues of $\mathbf{D} = \exp(\mathbf{B})$ for $L = 2$, $d_1 = e^{\zeta_1}$ and $d_2 = e^{\zeta_2}$, we have from Proposition 1:

$$0 \leq \zeta_L \leq \zeta_{L-1} \leq \dots \leq \zeta_2 \leq \zeta_1 \leq 1.$$

$\Rightarrow 1 \leq e^{\zeta_L} \leq e^{\zeta_{L-1}} \leq \dots \leq e^{\zeta_2} \leq e^{\zeta_1} \leq e$. Therefore, $d_1, d_2 \in [1, e]$, and

$$d_1 = e^{\zeta_1} = e^{\frac{1}{2} + \frac{1}{2} \sqrt{1 - \frac{4 \det(\mathbf{A})}{\text{tr}^2(\mathbf{A})}}},$$

$$d_2 = e^{\zeta_2} = e^{\frac{1}{2} - \frac{1}{2} \sqrt{1 - \frac{4 \det(\mathbf{A})}{\text{tr}^2(\mathbf{A})}}}.$$

Based on Corollary 3.23, $\Rightarrow d_1 \geq e^{1/2}$, and $d_2 \leq e^{1/2}$ for a white noise series, and $d_1 \leq e, d_2 \geq 1$ for a trend series.

3.5.5 The effect of the window length and rank of the Hankel matrix

Here, we consider different cases to evaluate the effect of window length on the $\text{tr}(\mathbf{A}_{L,N})$ and the eigenvalues $\zeta_i^{L,N}$. Based on Cauchy's interlacing theorem, the following theorem was given and proved in [217]; see also [218].

Theorem 3.25. *Let \mathbf{H} be an $L \times K$ Hankel matrix. Then, the eigenvalues of $\mathbf{H}\mathbf{H}^T$ have the following order:*

$$\gamma_j^{L,N} \geq \gamma_j^{L-l,N-l} \geq \gamma_{j+l}^{L,N}, \quad (j = 1, \dots, L-l),$$

where l is a number belonging to the set $\{1, \dots, L-1\}$.

Since $\zeta_j = \gamma_j / \text{tr}(\mathbf{A})$, we can write

$$\zeta_j^{L-l,N-l} \text{tr}(\mathbf{A}_{L-l,N-l}) \geq \zeta_{j+l}^{L,N} \text{tr}(\mathbf{A}_{L,N})$$

Note that $\zeta_j^{L-l,N-l}$ is the eigenvalue of the sub-matrix \mathbf{B}_1 , which is gained from a Hankel matrix corresponding to the sub-series $\mathbf{H}_{N-l} = (y_1, \dots, y_{N-l})$, where

$$\mathbf{B} = \begin{pmatrix} \mathbf{B}_1 & \mathbf{B}_2 \\ \mathbf{B}_3 & \mathbf{B}_4 \end{pmatrix}, \quad (3.43)$$

and

$$\mathbf{B}_1 = \frac{1}{\text{tr}(\mathbf{A})} \begin{pmatrix} \sum_{j=1}^K y_j^2 & \sum_{j=1}^K y_j y_{j+1} & \cdots & \sum_{j=1}^K y_j y_{j+L-l-1} \\ \sum_{j=1}^K y_j y_{j+1} & \sum_{j=1}^K y_{j+1}^2 & \cdots & \sum_{j=1}^K y_{j+1} y_{j+L-l-1} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{j=1}^K y_j y_{j+L-l-1} & \sum_{j=1}^K y_{j+1} y_{j+L-l-1} & \cdots & \sum_{j=1}^K y_{j+L-l-1}^2 \end{pmatrix}. \quad (3.44)$$

Our interest is to see the relationship between $\zeta_j^{L-l,N}$ and $\zeta_j^{L,N}$, not between $\zeta_j^{L-l,N-l}$ and $\zeta_j^{L,N}$. Thus, Theorem 3.25 should not be used directly.

In [217], it is mentioned that the behaviour of $\text{tr}(\mathbf{A}_{L,N})$ is similar on two intervals $2 \leq L \leq (N+1)/2$ and $(N+1)/2 + 1 \leq L \leq N-1$. They also showed that $\gamma_j^{L,N}$ increases as L increases until $L = (N+1)/2$, and decreases for the interval $(N+1)/2 + 1 \leq L \leq N-1$. Thus, it is enough to consider only one of these intervals. It is also enough to consider the behaviour of $\text{tr}(\mathbf{A}_{L,N})$ to observe the relationship between $\zeta_j^{L-l,N}$ and $\zeta_j^{L,N}$, since $\zeta_j^{L,N} = \gamma_j^{L,N} / \text{tr}(\mathbf{A}_{L,N})$.

Theorem 3.26. *Let $\text{tr}(\mathbf{A}_{L,N})$ be defined as in Eq. (3.3). Then, $\text{tr}(\mathbf{A}_{L,N})$ is an increasing function of L on the interval $\{2, \dots, (N+1)/2\}$, and a decreasing function of L on the interval $\{(N+1)/2 + 1, \dots, N-1\}$, and*

$$\max(\text{tr}(\mathbf{A}_{L,N})) = \text{tr}(\mathbf{A}_{(N+1)/2,N}).$$

Proof. Let us first show that $w_j^{L,N}$ is an increasing function of L on $\{2, \dots, (N+1)/2\}$. Assume L_1 and L_2 are two arbitrary values, and $L_1 < L_2 \leq (N+1)/2$. Then, from the definition of $w_j^{L,N}$ we obtain:

$$C_{j,L,N} = \begin{cases} 0, & \text{for } 1 \leq j \leq L_1, \\ j - L_1, & \text{for } L_1 + 1 \leq j \leq L_2, \\ L_2 - L_1, & \text{for } L_2 + 1 \leq j \leq N - L_2 + 1, \\ N - j + 1 - L_1 & \text{for } N - L_2 + 2 \leq j \leq N - L_1 + 1, \\ 0 & \text{for } N - L_1 + 2 \leq j \leq N \end{cases}$$

Thus, $w_j^{L_1,N} - w_j^{L_2,N} \geq 0$ for all j , and inequality is strict for some j [217]. Therefore,

$$tr(\mathbf{A}_{L_2,N}) - tr(\mathbf{A}_{L_1,N}) = \sum_{j=1}^N (w_j^{L_2,N} - w_j^{L_1,N}) y_j^2 > 0. \quad (3.45)$$

Eq. (3.45) confirms that $tr(\mathbf{A}_{L,N})$ is an increasing function of L on $\{2, \dots, (N+1)/2\}$. We use the same approach for $\{(N+1)/2 + 1, \dots, N-1\}$, which indicates that $tr(\mathbf{A}_{L,N})$ is a decreasing function of L on this interval. It should be noticed that increasing the value of L_2 leads to the increase of $tr(\mathbf{A}_{L_2,N}) - tr(\mathbf{A}_{L_1,N})$ in Eq. (3.45), which proves that $tr(\mathbf{A}_{L,N})$ is an increasing function on $\{2, \dots, (N+1)/2\}$. Consequently, the maximum value of $tr(\mathbf{A}_{L_2,N})$ is obtained at $L = (N+1)/2$, which is the maximum value of L . \square

Now, we consider four cases with different values of L and rank of \mathbf{H} to observe the relationship between $\zeta_j^{L,N}$ and $\zeta_j^{L-l,N}$.

Case 1 $L \geq 1$, rank = 1. In this case, we have only $\zeta_1^{L,N} = \frac{\gamma_1^{L,N}}{tr(\mathbf{A}_{L,N})} = 1$. Hence, $\zeta_1^{L,N} = \zeta_1^{L-l,N}$ for $2 \leq L \leq N-1$.

Case 2 $L = 1, 2$, rank = 2. We have $\zeta_1^{1,N}$, $\zeta_1^{2,N}$ and $\zeta_2^{2,N}$. If $L = 1$, then we only have one eigenvalue $\zeta_1^{1,N}$. If $L = 2$, then we have $\zeta_1^{2,N}$ and $\zeta_2^{2,N}$, which are defined in Eqs. (3.40) and (3.41). Furthermore, we know that $\zeta_1^{2,N} < 1$, $\zeta_1^{2,N} \geq 1/L$, and $\zeta_2^{2,N} = 1 - \zeta_1^{2,N}$.

$$\Rightarrow \zeta_1^{2,N} < 1 = \zeta_1^{1,N} \Rightarrow \zeta_1^{2,N} < \zeta_1^{1,N}.$$

Case 3 $L > 2$ and rank = 2. In this case, we use the the characteristic polynomial given by Eq. (3.34) to find the eigenvalues. As the rank of \mathbf{H} is two, then we

have $\zeta_{i=1,2}^{L,N} \neq 0$, which were given by Eqs. (3.40) and (3.41), and then $\zeta_i^{L,N} = 0$ ($i = 3, \dots, L$). In this regard, this case is similar to Case 2. The result shows that $\zeta_1^{L,N}$ decreases with L till $(N+1)/2$, then increases. In terms of $\zeta_2^{L,N}$ behaviour, it should be noted that as $\zeta_1^{L,N}$ decreases, $\zeta_2^{L,N}$ increases because $\zeta_2^{L,N} = 1 - \zeta_1^{L,N}$. Consequently, $\zeta_2^{L,N} \geq \zeta_2^{L-l,N}$ for $l = 1, \dots, (N+1)/2$, and $\zeta_2^{L,N} \leq \zeta_2^{L-l,N}$ for $l = (N+1)/2, \dots, N-1$.

Case 4 $L > 2$ and $\text{rank} > 2$. Using Eq. (3.35), we obtain the characteristic equation whose solution provides the eigenvalues of \mathbf{B} . However, their functional forms are very complicated in this case. Therefore, we consider several series to show the effect of L on the eigenvalues $\zeta_i^{L,N}$ in the following section.

3.5.6 Examples for evaluation of the theoretical results

Here, we give some examples related to some of the theoretical results obtained in the previous section, particularly the behaviour of $I_{\mathbf{A}}$ for different sample size, N , and the effect of window length and rank of \mathbf{H} on the eigenvalues of $\mathbf{B}_{L,N}$, in order to evaluate the results.

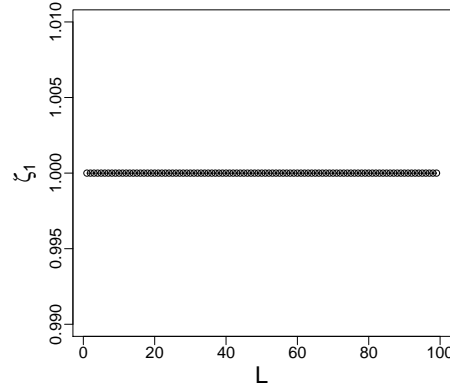
3.5.6.1 The effect of L on the eigenvalues

We present four examples to show the influence of L and rank of \mathbf{H} on the eigenvalues.

Examples

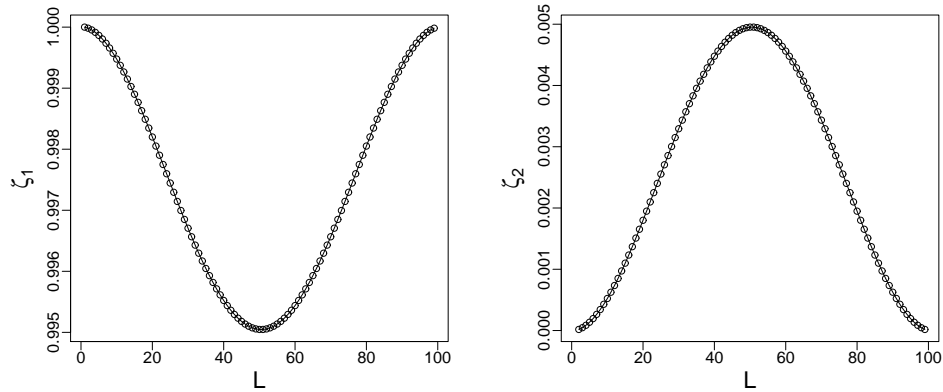
Example 1. Let $y_t = \exp(\alpha + \beta t)$, where $t = 1, \dots, N$, $N = 100$, $\alpha = 0.1$ and $\beta = 0.2$. In this case, the corresponding Hankel matrix \mathbf{H} has rank one [6, 32]. Figure. 3.3 illustrates $\zeta_1^{L,100}$ of $\mathbf{B}_{L,100}$ for this model, which is stable with respect to L and is equal to one.

In the next two examples, we consider two cases; the first is a linear model and the second is a sine model. The corresponding Hankel matrices in these examples have

FIGURE 3.3: Plot of ζ_1 for different values of L . Example 1.

rank two. It is easy to show that rank of the corresponding Hankel matrix \mathbf{H} is two; for more information refer to [6, 32]. Thus, we consider the behaviour of $\zeta_1^{L,N}$ and $\zeta_2^{L,N}$ with different values of L .

Example 2. Let $y_t = \alpha + \beta t$, where $t = 1, \dots, N$, $N = 100$, $\alpha = 1$ and $\beta = 2$. It can be seen from Figure. 3.4 that ζ_1 decreases with L until $L \leq (N+1)/2$, and then increases; it is convex and reaches minimum value at the median of $\{1, \dots, 100\}$. However, ζ_2 increases for $2 \leq L \leq (N+1)/2$ and decreases for $(N+1)/2 \leq L \leq N-1$, which is expected because if ζ_1 decreases, the second eigenvalue ζ_2 will increase as the sum of these two eigenvalues is equal to one.

FIGURE 3.4: Plot of ζ_1 (left) and ζ_2 (right) for different values of L . Example 2.

Example 3. Let $y_t = \sin(\pi t/12)$. Figure. 3.5 depicts ζ_1 and ζ_2 for this series with length 100. It is obvious that, if we ignore the small fluctuations in the plots, then the behaviour of the first and second eigenvalues is similar to the previous example. Roughly speaking, the results of this example and second one are similar.

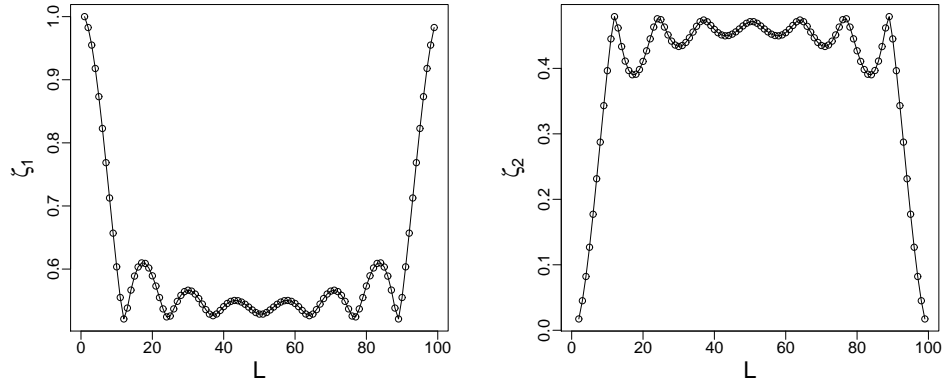


FIGURE 3.5: Plot of ζ_1 (left) and ζ_2 (right) for different values of L . Example 3.

Example 4. Let $y_t = \log(t)$, where $t = 1, \dots, N$, and $N = 100$. The rank of the corresponding Hankel matrix \mathbf{H} of this series is four [217]. The eigenvalues, $\zeta_i^{L,100}$ ($i = 1, \dots, 4$) are shown in Figure. 3.6. It can be seen that the results are also in concordance with all the previous examples given in this section, that the largest eigenvalue $\zeta_1^{L,N}$ is a decreasing function for $2 \leq L \leq (N+1)/2$, and an increasing function for $(N+1)/2 \leq L \leq N-1$, whilst the other eigenvalues ζ_i ($i = 2, \dots, 4$) increase with L up to the median of $\{1, \dots, 100\}$, then decrease. The

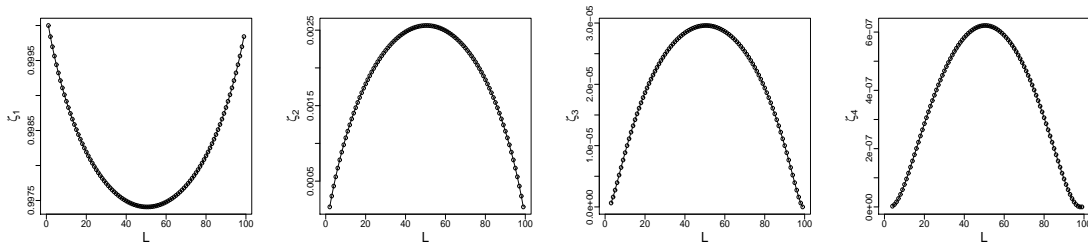


FIGURE 3.6: Plot of ζ_i ($i = 1, \dots, 4$) for different values of L . Example 4.

results obtained here agree with those results proved in the previous section.

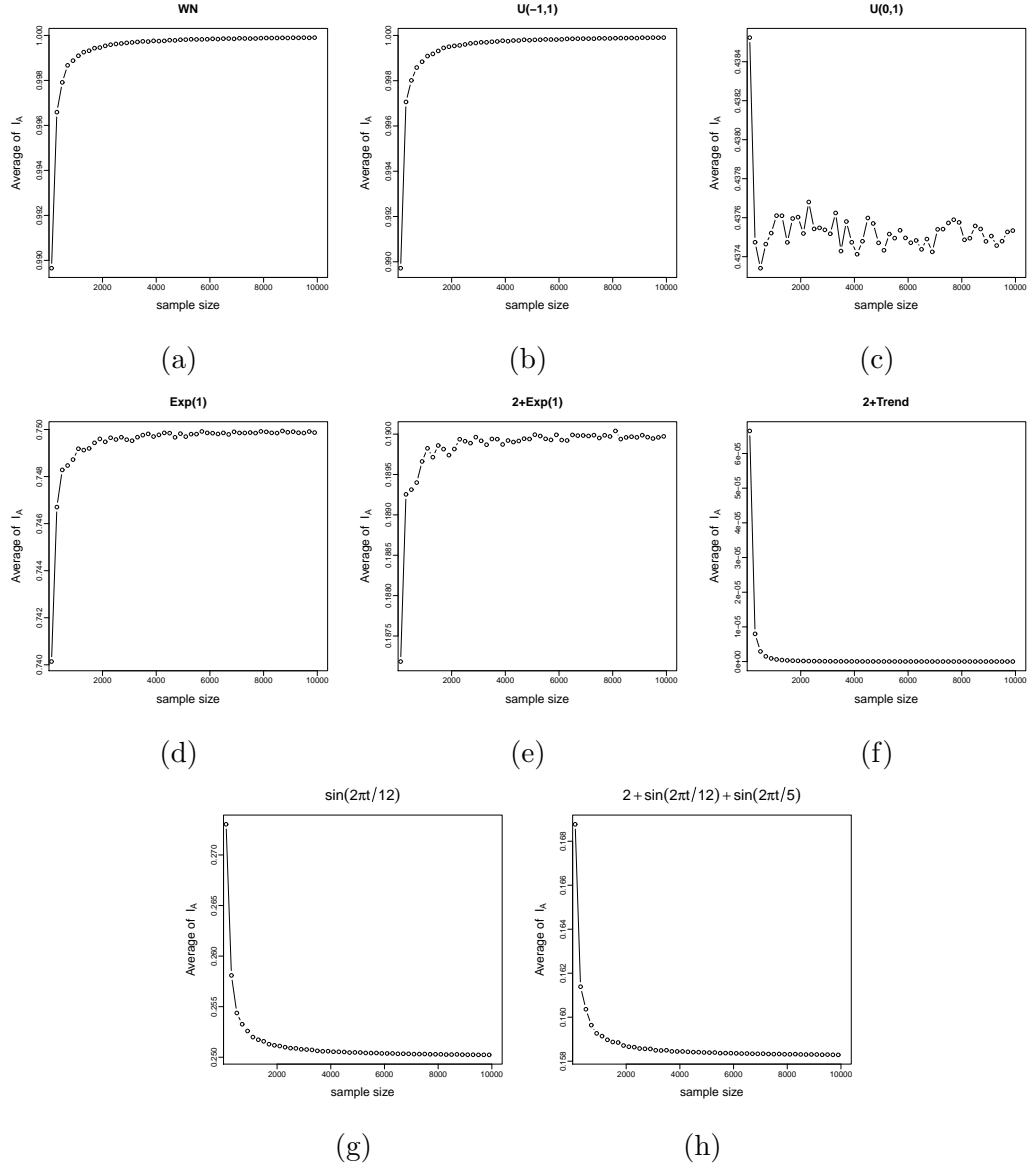
3.5.6.2 Asymptotical behaviour of $I_{\mathbf{A}}$ and ζ_i

Let us now consider the asymptotical behaviour of $I_{\mathbf{A}}$ defined in Eq. (3.42) for different sample size N (for simplicity, $L = 2$ is used here). The simulation was repeated 10^4 times. Figure. 3.7 shows the average of $I_{\mathbf{A}}$, $\bar{I}_{\mathbf{A}}$, for different values of N , for symmetric and nonsymmetric distributions, trend series, and sine wave. Here we consider eight different cases that can be seen in real life examples:

- (a) White noise; WN.
- (b) Uniform distribution with mean zero; $U(-\alpha, \alpha)$.
- (c) Uniform distribution; $U(0, \alpha)$.
- (d) Exponential distribution, $Exp(\alpha)$.
- (e) $\beta + Exp(\alpha)$.
- (f) $\beta + t$.
- (g) Sine wave series; $\sin(\varphi)$.
- (h) $\beta + \sin(\varphi) + \sin(\vartheta)$,

where, $\alpha = 1$, $\beta = 2$, $\varphi = 2\pi t/12$, $\vartheta = 2\pi t/5$, and t is the time which is used to generate the linear trend series. Note that if Y_N is the trend series, then the rank of $\mathbf{A}/tr(\mathbf{A})$ is one. Thus, $I_{\mathbf{A}}$ is approximately equal to zero based on Theorem 3.17.

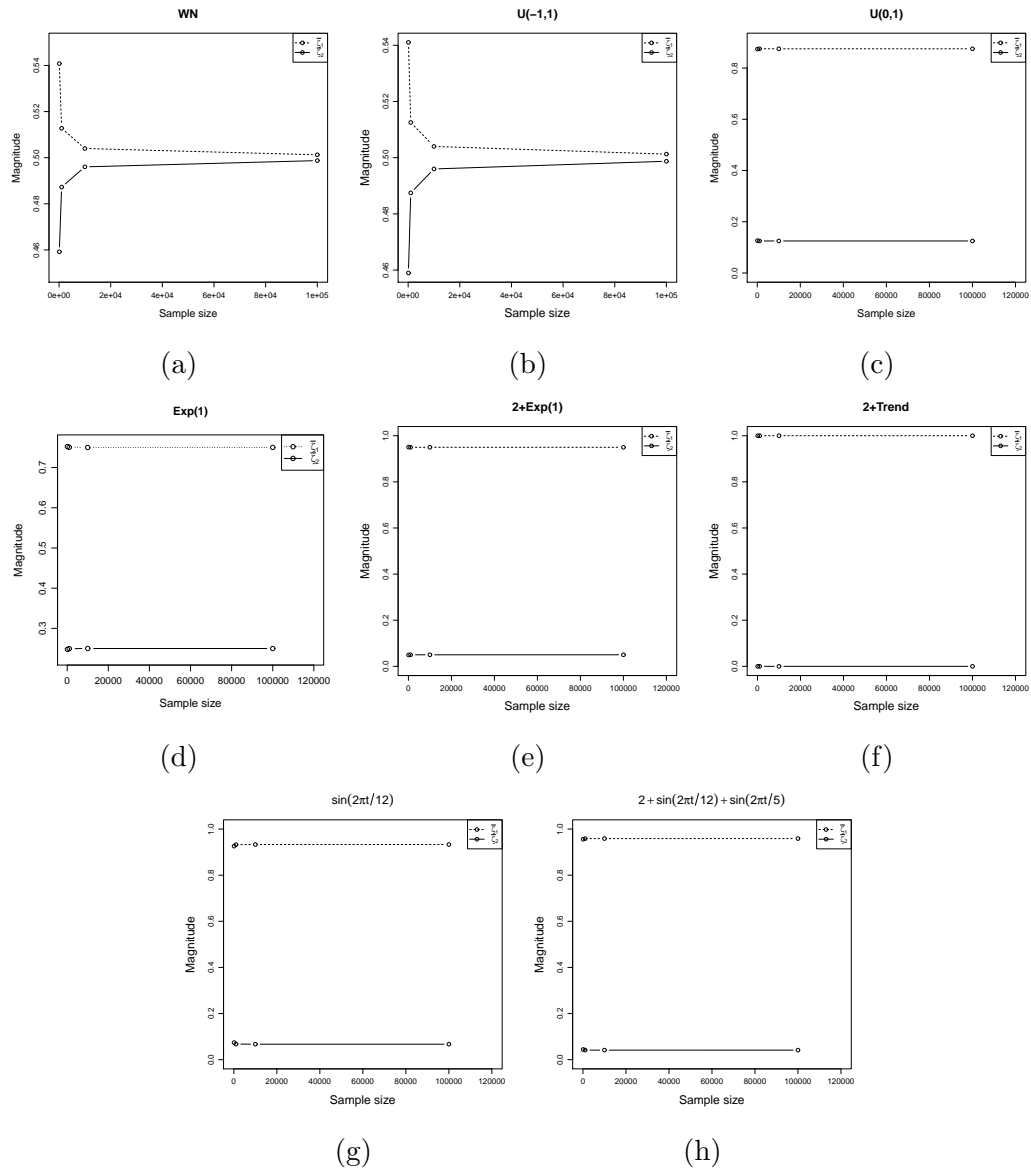
For cases (a) and (b), where Y_N is generated from a symmetric distribution, $\bar{I}_{\mathbf{A}} \approx 1$, then $\bar{\zeta}_{i=1,2} \approx 1/2$. However, from the second and third properties of Proposition 1, we have $\zeta_1 \geq 1/2$ and $\zeta_2 \leq 1/2$. Thus, $\zeta_1 \gtrsim 1/2$ and $\zeta_2 \lesssim 1/2$. It is obvious that $\bar{I}_{\mathbf{A}}$ converges asymptotically to $3/4$ for the exponential distribution (d) (see Figure. 3.7 (d)). Note also that adding a constant to the exponential (case (d)), will change the value of $\bar{I}_{\mathbf{A}}$ (see Figure. 3.7 (e)). For cases (f) and (g), the values of $\bar{I}_{\mathbf{A}}$, for a large N , are approximately $1/4$ and $4/25$ respectively (see Figure. 3.7 (f) and (g)). The result for the trend series is shown in Figure. 3.7 (f). It is clear

FIGURE 3.7: Plot of \bar{I}_A for different sample size N .

that $\bar{I}_A \approx \text{zero}$, which is supported by Theorem 3.17. Thus, $\zeta_1 \approx 1$ and $\zeta_2 \approx 0$. Note that the properties of Proportion 1 indicate that $\zeta_1 \gtrsim 1$ and $\zeta_2 \gtrsim 0$.

Let us now assess the asymptotical behaviour of ζ_1 and ζ_2 , for different values of N . Figure. 3.8 displays the results for $m = 5 \times 10^3$ simulations for all cases, where $\bar{\zeta}_i = \left(\sum_{j=1}^m \zeta_{i,j} \right) / m$, $i = 1, 2$. As the results indicate, ζ_1 and ζ_2 have asymptotic behaviour for the white noise (a) and uniform (b) distributions. The gap between ζ_1 and ζ_2 (or their averages after simulations, $\bar{\zeta}_1$, and $\bar{\zeta}_2$, respectively) become smaller as the sample size increases, and both converge to $\frac{1}{2}$. This result can be

clearly seen in Figure. 3.8 (a) and (b). Note that for any value of L , both ζ_1 and ζ_L converge to $\frac{1}{L}$. Thus, according to property 5 of Proposition 1, other eigenvalues tend to $\frac{1}{L}$. However, we have not observed the behaviour of $\bar{\zeta}_1$ and $\bar{\zeta}_2$ for other cases ($\bar{\zeta}_1$ and $\bar{\zeta}_2$ do not approach each other). Furthermore, it seems that ζ_i ($i = 1, 2$) has a stable pattern for different values of N (see Figure. 3.8 (c),..., (h)). Note that the convergence between ζ_1 and ζ_L may tend to $1/L$ depending on both the distribution of Y_N and sample size N . Moreover, the rate $I_{\mathbf{A}}$ plays a significant role in identifying the value of ζ_i . For example, if $Y_N \sim \text{Exp}(1)$, then $\bar{I}_{\mathbf{A}} \approx 3/4$ (see Figure. 3.7 (d)). Thus, $\bar{\zeta}_1 \approx 17/20$, and $\bar{\zeta}_2 \approx 3/20$ (see Figure. 3.8 (d)). This conclusion can be extended to other cases.

FIGURE 3.8: Plot of $\bar{\zeta}_1$, and $\bar{\zeta}_2$ for different sample size N .

Chapter 4

A Study on the Distribution of the Scaled Hankel Matrices Eigenvalues

4.1 Introduction

In this chapter, the main task is to evaluate the empirical distribution of ζ_i . In doing so, a series of length N from different distributions, is generated m times. We consider the same eight cases used in the previous chapter. The results indicate that even adding an intercept alone will change the distribution of ζ_i . Note that an intercept can be considered as a trend in time series analysis. Note also that usually every harmonic component with a different frequency produces two close eigenvalues (except for frequency 0.5 which provides one eigenvalue). It will be clearer if N , L , and K are sufficiently large [112]. In practice, the eigenvalues of a harmonic series are often close to each other, and this fact simplifies the visual identification of the harmonic components [112].

Generally, the results confirm that if we add more non stochastic components to the noise series, for instance trend, harmonic and cyclical components, then the first few eigenvalues are related to those components and as soon as we reach the noise

level the pattern of eigenvalues will be similar to those found for the noise series. Thus, the results obtained here are very interesting for signal processing and time series techniques where noise reduction and filtering matter, and the results are very important for the following chapters.

The chapter also covers the effect of N on the mean and variance of ζ_i ; the patterns of the mean; variance, and coefficient of variations for each eigenvalue for a fixed $N = 10^5$; the empirical distribution of each eigenvalue; the effect of the mean and variance of Gaussian distribution on the eigenvalues; and some examples related to the effect of window length on the distribution of largest and smallest eigenvalues for white noise process.

For consistency and comparability of the results, a fixed value of L , here 10, is used for all examples and case studies throughout the chapter, except in section 4.7. For point estimation and comparison of the mean value of the eigenvalues, the average of each eigenvalue in m runs is used; $\bar{\zeta}_i$ as defined before, $i = 1, \dots, L$. Similarly, the variance and the coefficient of variation of ζ_i are also evaluated: $\sigma_i^2 = \left(\sum_{n=1}^m (\zeta_{i,n} - \bar{\zeta}_i)^2 \right) / m$, and $CV_i = \frac{\sqrt{\sigma_i^2}}{\bar{\zeta}_i}$, respectively.

4.2 The effect of N

In this section, we consider the effect of N on $\bar{\zeta}_i$ and σ_i^2 . Figure 4.1 demonstrates $\bar{\zeta}_i$ for different values of N for all cases ((a), ..., (h)) considered in this chapter. In Figure 4.1, $\bar{\zeta}_i$ has a decreasing pattern for different values of N . It can be seen that, for a large N , $\bar{\zeta}_i \rightarrow 1/10$ for cases (a) and (b). Thus, increasing N clearly affects the values of $\bar{\zeta}_i$ for the white noise (a) and uniform distribution (b). However, there is no obvious effect on $\bar{\zeta}_i$ for other cases. For example, for case (c), $\bar{\zeta}_1$ is approximately equal to 0.8 for different values of N , and $\bar{\zeta}_{i \neq 1}$ is less than 1/10 (see Figure 4.1 (c)).

Although the pattern of $\bar{\zeta}_i$ for uniform distribution (c) is similar to exponential case (d), for case (c), $\bar{\zeta}_1$ is greater than $\bar{\zeta}_1$ compared to case (d), whilst other $\bar{\zeta}_i$ are smaller. It has been observed that $\bar{\zeta}_i$ has similar patterns for cases ((c), ..., (f)).

The values of $\bar{\zeta}_i$ for cases (a) and (b), where Y_N is generated from a symmetric distribution, are approximately the same. The results clearly indicate that increasing N does not have a significant influence on the mean of ζ_i in all cases, except (a) and (b). As a result, if Y_N is generated from WN or $U(-1, 1)$, then increasing N will affect the value of $\bar{\zeta}_i$ significantly.

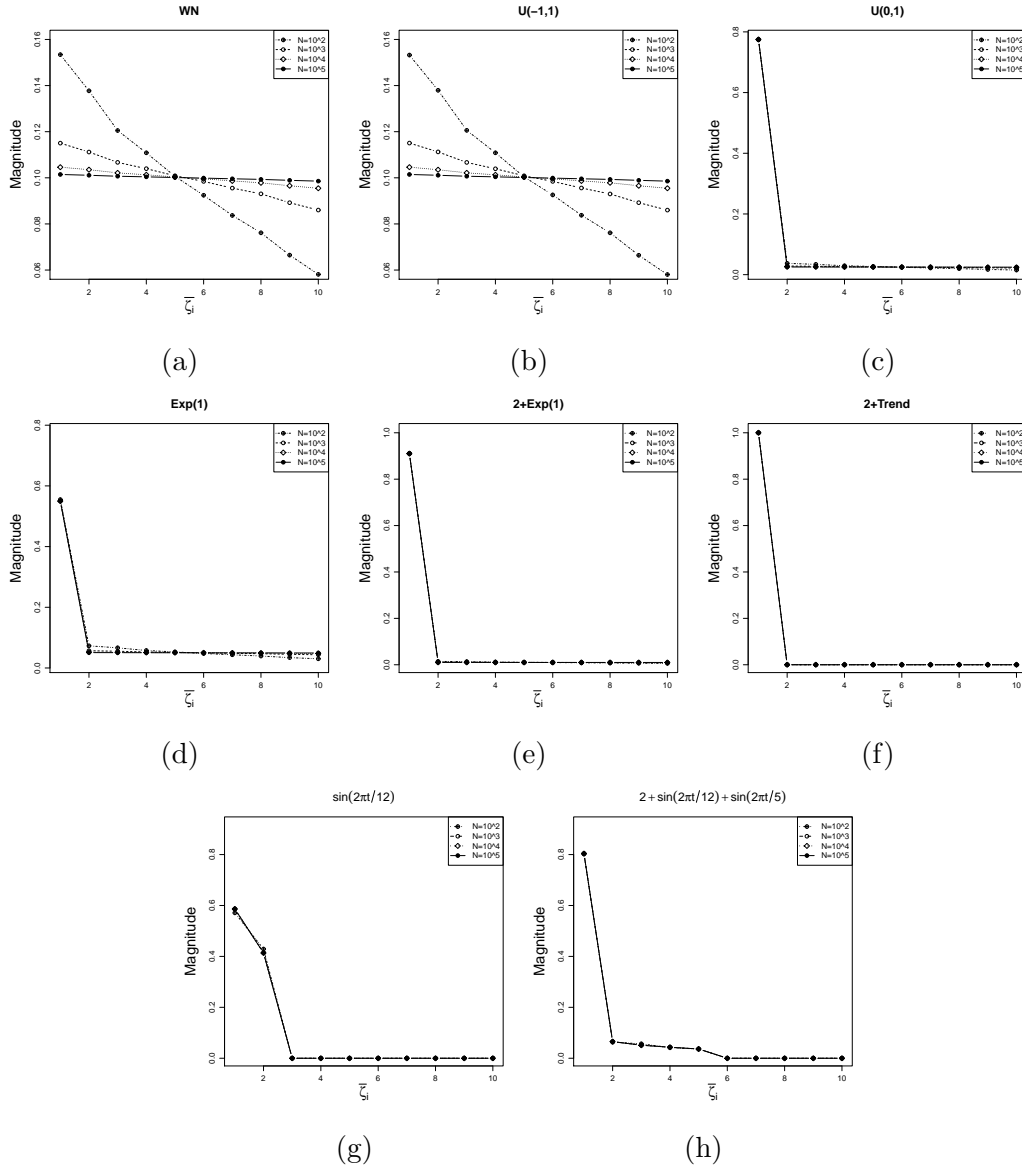


FIGURE 4.1: Plot of $\bar{\zeta}_i$, ($i = 1, \dots, 10$) for different values of N .

In terms of σ_i^2 for the WN process, Figure. 4.2 demonstrates σ_i^2 for different values of N . It is clear that the minimum value of σ_i^2 is in the middle. Generally, for a large N and small L , $\sigma_1^2 \simeq \sigma_L^2$. Thus, σ_i^2 has a U shape with a global minimum.

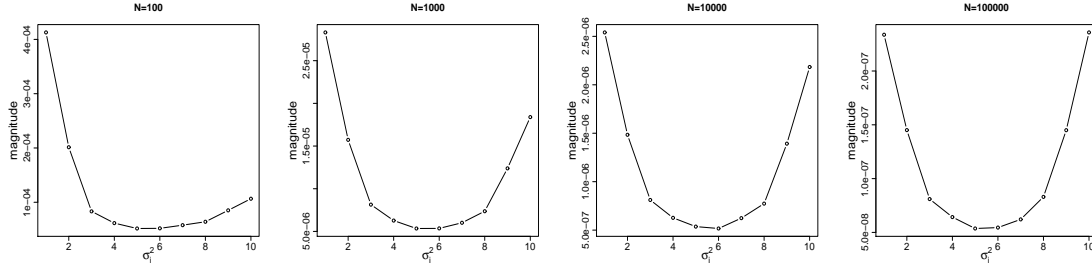
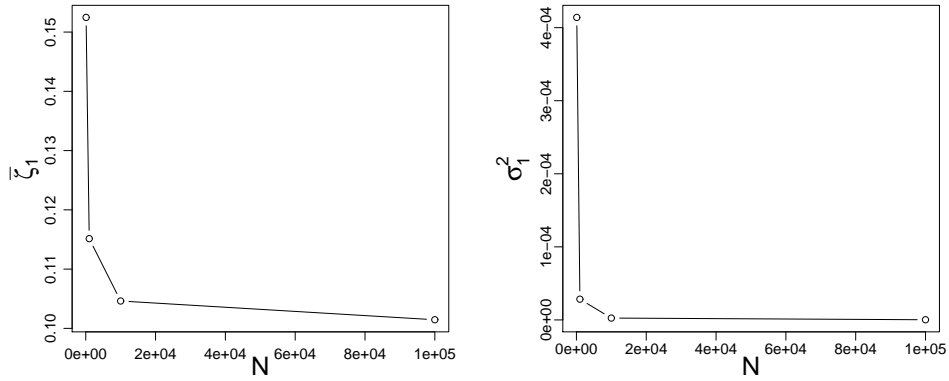
FIGURE 4.2: Plot of σ_i^2 for different values of N for WN.

Figure. 4.3 also shows the pattern of $\bar{\zeta}_1$ (left) and σ_1^2 (right) for different values of N for WN. The pattern of $\bar{\zeta}_1$ indicates clearly that $\bar{\zeta}_1$ has a decreasing order for different values of N . Similar results are also obtained for σ_1^2 (see Figure. 4.3 (right)). In addition to the previous results, the coefficient of variation CV_i of ζ_i for

FIGURE 4.3: The plot of $\bar{\zeta}_1$ (left) and σ_1^2 (right) for different values of N for WN.

WN is also evaluated. The patterns of CV_i are similar to what emerged for σ_i^2 .

4.3 The patterns of $\bar{\zeta}_i$

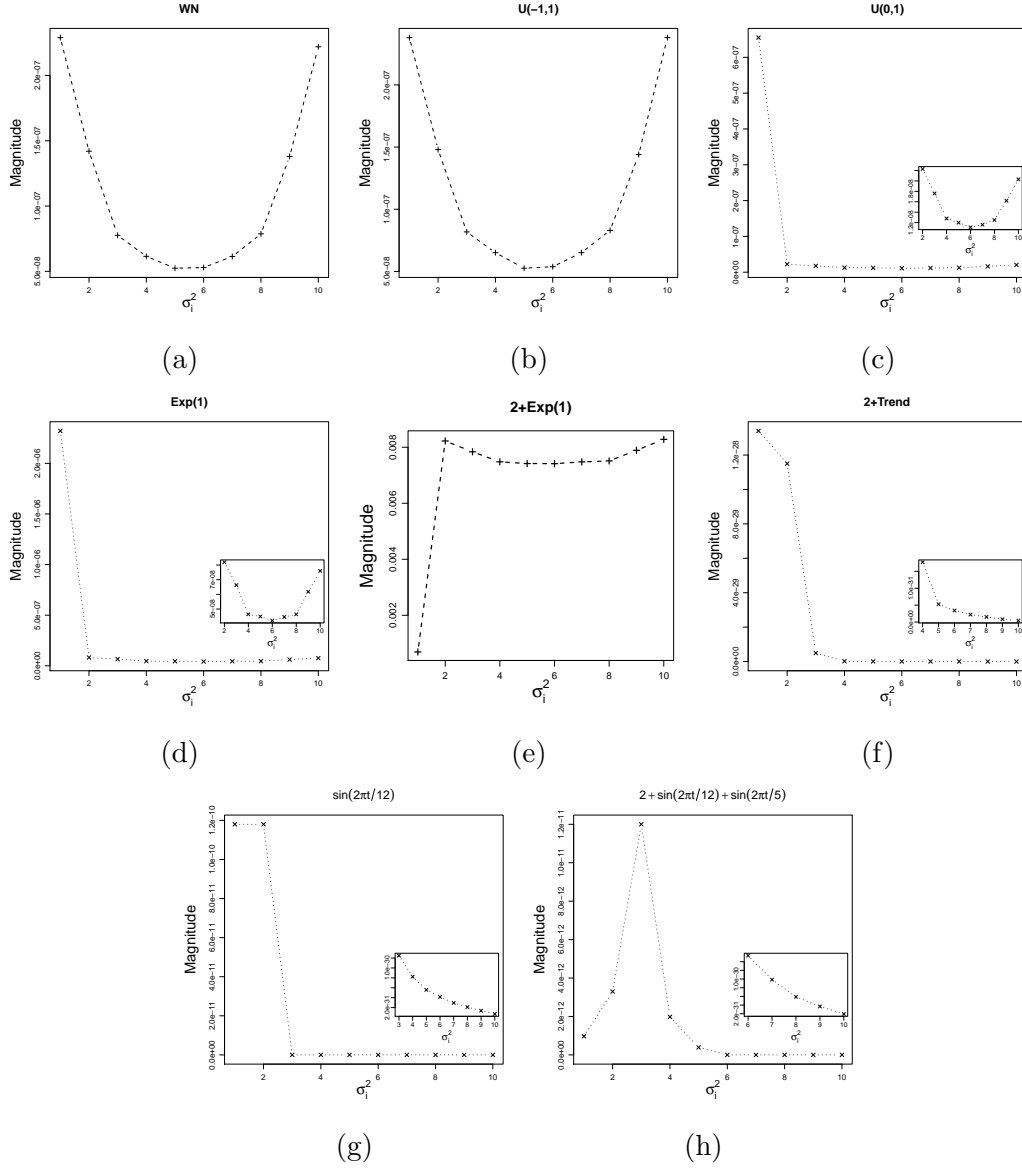
Let us now focus only on the patterns of $\bar{\zeta}_i$ for $N = 10^5$. For the white noise distribution (a) and trend series (f), $\bar{\zeta}_i$ has different patterns. It is obvious that for the white noise series, $\bar{\zeta}_i$ converges asymptotically to $1/10$, whilst for the trend series $\bar{\zeta}_1$ is approximately equal to 1, and $\bar{\zeta}_{i \neq 1}$ tends to zero. Similar results were obtained for the uniform distributions, cases (b) and (c), respectively.

Both samples generated from exponential distribution have similar patterns for $\bar{\zeta}_i$. However, it was noticed that adding an intercept β to the exponential distribution increases the value of $\bar{\zeta}_1$ and decreases other $\bar{\zeta}_i$. The results indicate that $\bar{\zeta}_1 \approx 0.6$ and $\bar{\zeta}_2 \approx 0.4$, whilst other $\bar{\zeta}_i \approx \text{zero}$ for sine wave (g). It also indicates that for sine case (h) $\bar{\zeta}_i$ ($i = 1, \dots, 5$) is not zero, whereas other $\bar{\zeta}_i$ tend to zero. It was also noticed that the value of $\bar{\zeta}_1$ for sine wave (h) is greater than its value for sine case (g), whilst the value of $\bar{\zeta}_2$ is lower.

4.4 The patterns of σ_i^2 and CV_i^2

Figure. 4.4 shows the results of σ_i^2 for all cases (a), ..., (h) considered in Chapter 3. It can be seen that $\sigma_1^2 \simeq \sigma_L^2$ for cases (a) and (b). Consequently, σ_i^2 has a *U* shape with a global minimum for the white noise and uniform distribution with zero mean. However, the patterns of σ_i^2 for cases (c) and (d) are different to what was observed for cases (a) and (b). Note that σ_i^2 ($i = 2, \dots, 10$) are not stable for the uniform (c) and exponential (d) distributions. For example, if we remove σ_1^2 , the pattern becomes similar to what was seen for the white noise series and the sample taken from the uniform distribution (b) (see Figure. 4.4 (c) and (d)). For the sine wave (g), σ_i^2 has a *Z* shape. Note that σ_i^2 ($i = 3, \dots, 10$) are also not stable and have a decreasing order in this case. Similar results were observed for case (f), where Y_N is a trend. Looking at σ_i^2 for case (h), σ_i^2 increases for ($i = 1, 2, 3$), then decreases. It should be noted that σ_i^2 ($i = 6, \dots, 10$) has a decreasing order (see Figure. 4.4 (h)).

In addition to the previous results, the coefficient of variation CV_i of ζ_i is also assessed. The pattern of CV_i for the white noise and uniform distribution (b) are similar to what was observed earlier for σ_i^2 (see Figure. 4.5 (a) and (b)). It can also be seen that CV_i have similar patterns for cases (c) and (d). Note that the value of CV_1 for exponential case (d) is greater than its value for exponential case (e) (with an intercept). For both sine series, CV_i has similar behaviour. It can be seen that $CV_i \approx \text{zero}$ ($i = 1, 2$) for sine wave (g), and $CV_i \approx \text{zero}$ ($i = 1, \dots, 5$) for

FIGURE 4.4: Plot of σ_i^2 .

sine wave (h), whilst other CV_i have an increasing order for both cases (see Figure 4.5 (g) and (h)). For the trend case, $CV_i \approx \text{zero}$ ($i = 1, 2$), whilst other CV_i have an approximate U shape.

4.5 The empirical distribution of ζ_i

The distribution of ζ_i was assessed for all cases. The results are provided only for ζ_1 , ζ_5 and ζ_{10} , for cases ((a), ..., (d)), as similar results are observed for other ζ_i .

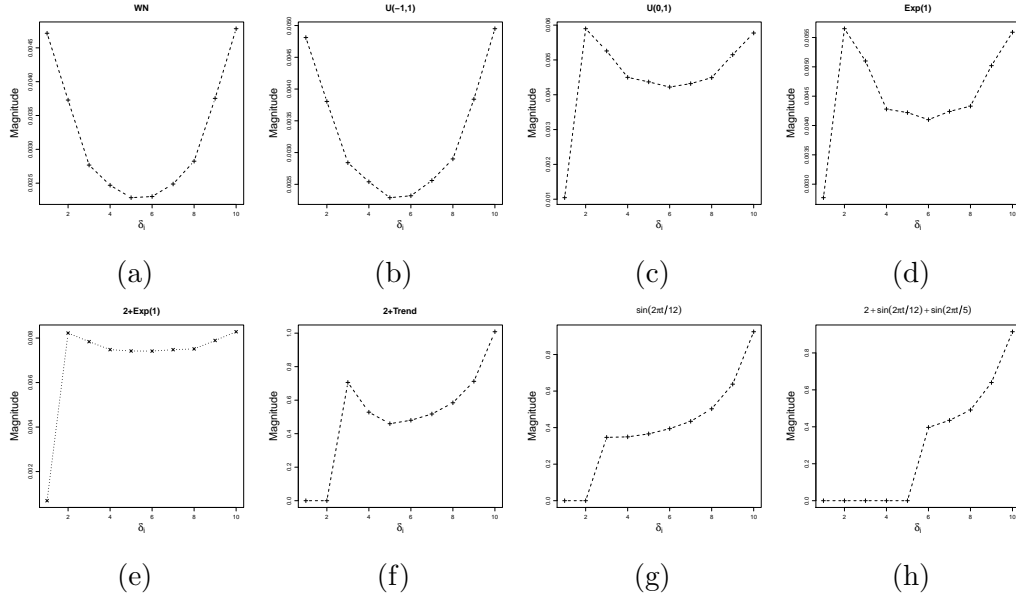
FIGURE 4.5: Plot of CV_i .

Figure 4.6 shows the histogram of ζ_i ($i = 1, 5, 10$) for $L = 10$, and $m = 5 \times 10^3$ simulations. It appears that the histogram of ζ_1 , is skewed to the right for samples taken from WN (a) and uniform distributions (b), whilst for the data generated from the uniform (c) and exponential (d) distributions, it possibly be symmetric. For the middle ζ_i , the histogram might be symmetric for the four cases (results are only provided for ζ_5), whilst the distribution of ζ_{10} , is skewed to the left.

For the cases of exponential distribution (e), trend series (f), sine wave series (g) and complex series (h), we have standardised ζ_i to convey information about their distributions. Figure 4.7 shows the density of ζ_i ($i = 1, 2, 3, 5, 6, 10$) for those cases. It is clear that ζ_1 has a different histogram for these cases, and is also different from what was achieved for the white noise and uniform distributions with zero mean. Note that if Y_N is generated from a symmetric distribution, as in cases (a) and (b), ζ_1 has a right skewed distribution. Moreover, it is interesting that ζ_{10} has a negative skewed distribution for all cases except the trend series and sine ones ((g), (h)).

Additionally, it should be noted that, for sine series (g), both ζ_1 and ζ_2 have similar distributions, whereas other ζ_i have right skewed distributions. It is obvious that the distribution of ζ_i for sine series (h) becomes skewed to the right for ζ_i ($i = 6, \dots, 10$). Remember that sine wave (h) was generated from an intercept and two pure sine

waves. This means that the components related to the first five eigenvalues create sine series (h).

In terms of the distribution of ζ_i for the trend series and sine wave (g), the distributions of $\zeta_{i=1,2}$ are totally different to the distributions of other ζ_i , which become skewed. Note that the distribution of ζ_i ($i = 1, 2$) for the trend series is symmetric, whilst skewed for the sine wave (g). For sine series (h), the distribution of ζ_i ($i = 1, \dots, 5$) is different from the distribution of ζ_i ($i = 6, \dots, 10$). It is obvious from Figure 4.7 that ζ_i ($i = 6, 10$) has a right skewed distribution.

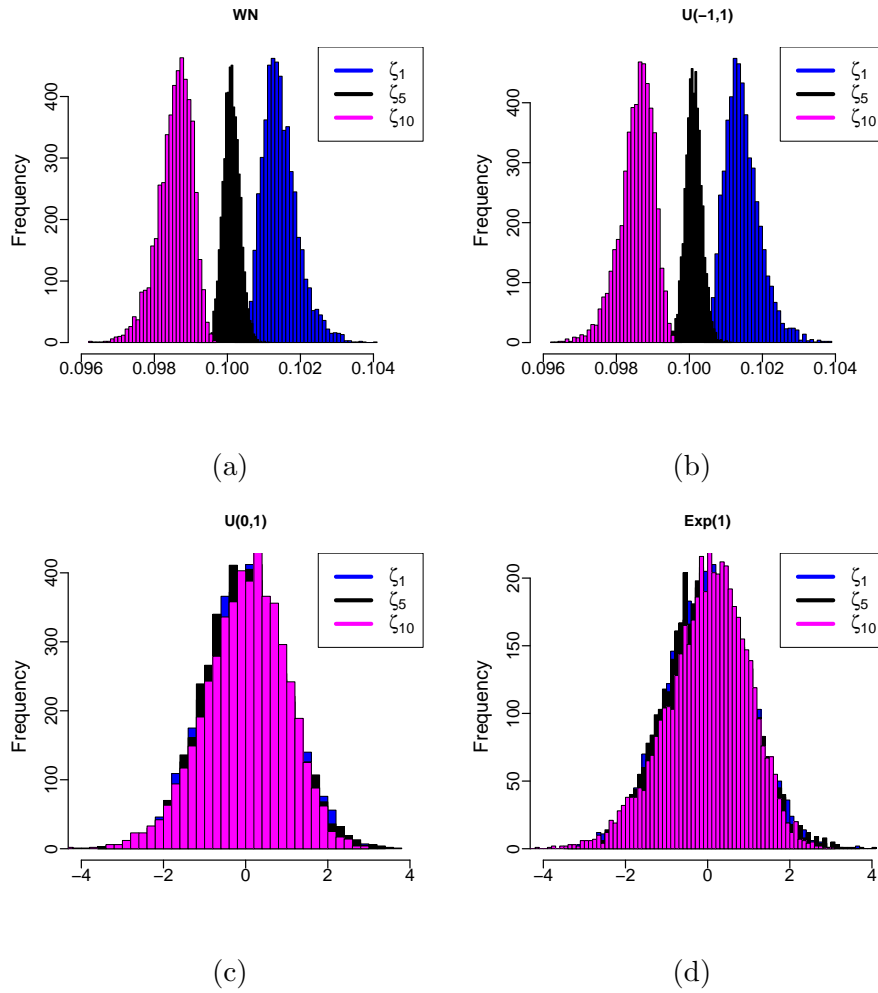
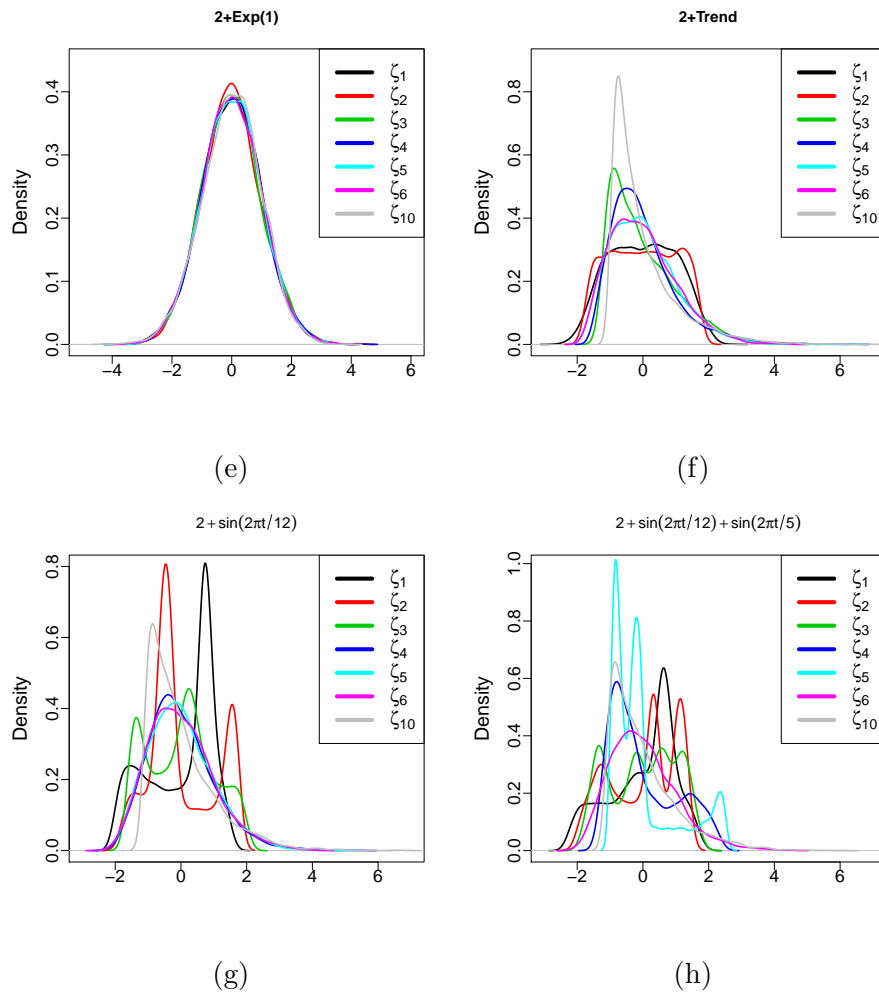


FIGURE 4.6: Histograms of ζ_1 , ζ_5 , and ζ_{10} for cases ((a), ..., (d)).

Generally, it is not easy to judge visually if ζ_i has a symmetric distribution, thus it is necessary to consider other criteria such as statistical tests. We calculate the coefficient of skewness, which is a measure for the degree of symmetry in the

FIGURE 4.7: Density of ζ_i , $i = 1, \dots, 6, 10$ for cases ((e), ..., (h)).

distribution of a variable. Table 4.1 represents the coefficient of skewness of ζ_i for all cases. Bulmer [219] suggests that if skewness is less than -1 or greater than +1, the distribution is highly skewed; if it is between -1 and -1/2 or between +1/2 and +1, the distribution is moderately skewed; and finally if it is between -1/2 and +1/2, the distribution is approximately symmetric. Therefore, we can say that, for instance, the distribution of ζ_1 for cases ((e), ..., (h)), and ζ_5 for all cases might be symmetric.

The D'Agostino-Pearson normality test (D-P) [220] (see also Appendix B), is applied here to evaluate this issue properly (see Table 4.2). It is also known as the omnibus test because it uses the test statistics for both the skewness and kurtosis to come up with a single p -value and to quantify how far from Gaussian the distribution is in terms of asymmetry and shape. The p -value of the D-P test was not significant;

Coefficient of Skewness of $\zeta_i, i = 1, \dots, 10$.								
	WN	$U(-1, 1)$	$U(0, 1)$	$Exp(1)$	$2 + Exp(1)$	$2 + t$	$\sin(\varphi)$	$2 + \sin(\varphi) + \sin(\vartheta)$
ζ_1	0.991	0.934	-0.056	-0.021	0.023	-0.046	0.341	0.157
ζ_2	0.692	0.687	0.491	0.423	0.117	0.021	-0.341	-0.079
ζ_3	0.461	0.564	0.235	0.328	0.048	1.103	0.734	-0.195
ζ_4	0.401	0.352	0.078	0.065	0.019	1.405	0.631	-0.321
ζ_5	0.099	0.057	0.070	0.071	-0.026	0.794	0.678	-0.059
ζ_6	-0.140	-0.098	0.006	0.070	-0.031	0.828	0.708	0.754
ζ_7	-0.37	-0.442	0.057	0.001	-0.010	0.921	0.761	0.632
ζ_8	-0.503	-0.572	-0.003	-0.129	-0.030	1.123	0.930	0.744
ζ_9	-0.577	-0.617	-0.174	-0.278	-0.063	1.540	1.02	0.998
ζ_{10}	-0.810	-0.875	-0.568	-0.479	-0.187	2.058	1.616	1.516

TABLE 4.1: Coefficient of skewness for $\zeta_i, (i = 1, \dots, 10)$, for all cases.

greater than 0.05 for ζ_1 , for cases ((c), ..., (f)), whereas, it is less than 0.05 for other cases ((a), (b), (g), (h)). Therefore, we accept the null hypothesis that the data of ζ_1 for cases ((c), ..., (f)) are symmetric and as a result are not skewed. Moreover, ζ_5 has a symmetric distribution for all cases, except the trend series and sine wave (g). Note that the distribution of ζ_i ($i = 2, 3, 9$), for the exponential case (d) is skewed, whereas it is symmetric for the exponential case with intercept (e).

P -value of D-P test for $\zeta_i, i = 1, \dots, 10$.								
	WN	$U(-1, 1)$	$U(0, 1)$	$Exp(1)$	$2 + Exp(1)$	$2 + t$	$\sin(\varphi)$	$2 + \sin(\varphi) + \sin(\vartheta)$
ζ_1	<2.2e-16	<2.2e-16	0.760*	0.362*	0.735*	0.501*	9.2e-7	0.021
ζ_2	<2.2e-16	<2.2e-16	2.1e-9	1.4e-11	0.089*	0.761*	9.2e-7	0.242*
ζ_3	2.7e-14	3.1e-14	2.3e-6	4.8e-05	0.474*	<2.2e-16	<2.2e-16	0.004
ζ_4	1.02e-06	6.3e-09	0.338*	0.250*	0.768*	<2.2e-16	<2.2e-16	3.8e-6
ζ_5	0.271*	0.851*	0.296*	0.290*	0.699*	<2.2e-16	8.4e-16	0.386*
ζ_6	0.442*	0.075*	0.299*	0.920*	0.651*	<2.2e-16	<2.2e-16	<2.2e-16
ζ_7	2.9e-07	3.1e-06	0.990*	0.402*	0.881*	<2.2e-16	<2.2e-16	<2.2e-16
ζ_8	4.1e-13	6e-10	0.057*	0.968*	0.658*	<2.2e-16	<2.2e-16	<2.2e-16
ζ_9	<2.2e-16	<2.2e-16	5.7e-5	0.001	0.355*	<2.2e-16	<2.2e-16	<2.2e-16
ζ_{10}	<2.2e-16	<2.2e-16	1.7e-11	3.2e-09	0.006	<2.2e-16	<2.2e-16	<2.2e-16

Note :*, represents symmetry based on D-P test at $p = 0.05$.

TABLE 4.2: p -value of the D'Agostino-Pearson test for $\zeta_i, (i = 1, \dots, 10)$, for all cases.

In addition, the relationship between ζ_i and $\zeta_j, (i, j = 1, \dots, 10)$ is also studied. The Pearson's correlation coefficient, ρ , between two sets of data, for example, ζ_i and ζ_j with expected values $\bar{\zeta}_i$ and $\bar{\zeta}_j$ (respectively) is considered, and defined as follows:

$$\rho = cor(\zeta_i, \zeta_j) = \frac{\sum_{n=1}^m (\zeta_{i,n} - \bar{\zeta}_i)(\zeta_{j,n} - \bar{\zeta}_j)}{\sqrt{\sum_{n=1}^m (\zeta_{i,n} - \bar{\zeta}_i)^2} \sqrt{\sum_{n=1}^m (\zeta_{j,n} - \bar{\zeta}_j)^2}}$$

Figure. 4.8 shows the correlation matrix for ζ_i considering Pearson correlation, which is the most commonly reported, even for those data for which it is superficially not a good match. The correlation matrix for ζ_i is considered, in a 20-grade grey scale from white to black corresponding to the values of correlations from -1 to 1. The white colour indicates that ζ_i and ζ_j have perfect negative correlation, whilst a black value shows a perfect positive correlation.

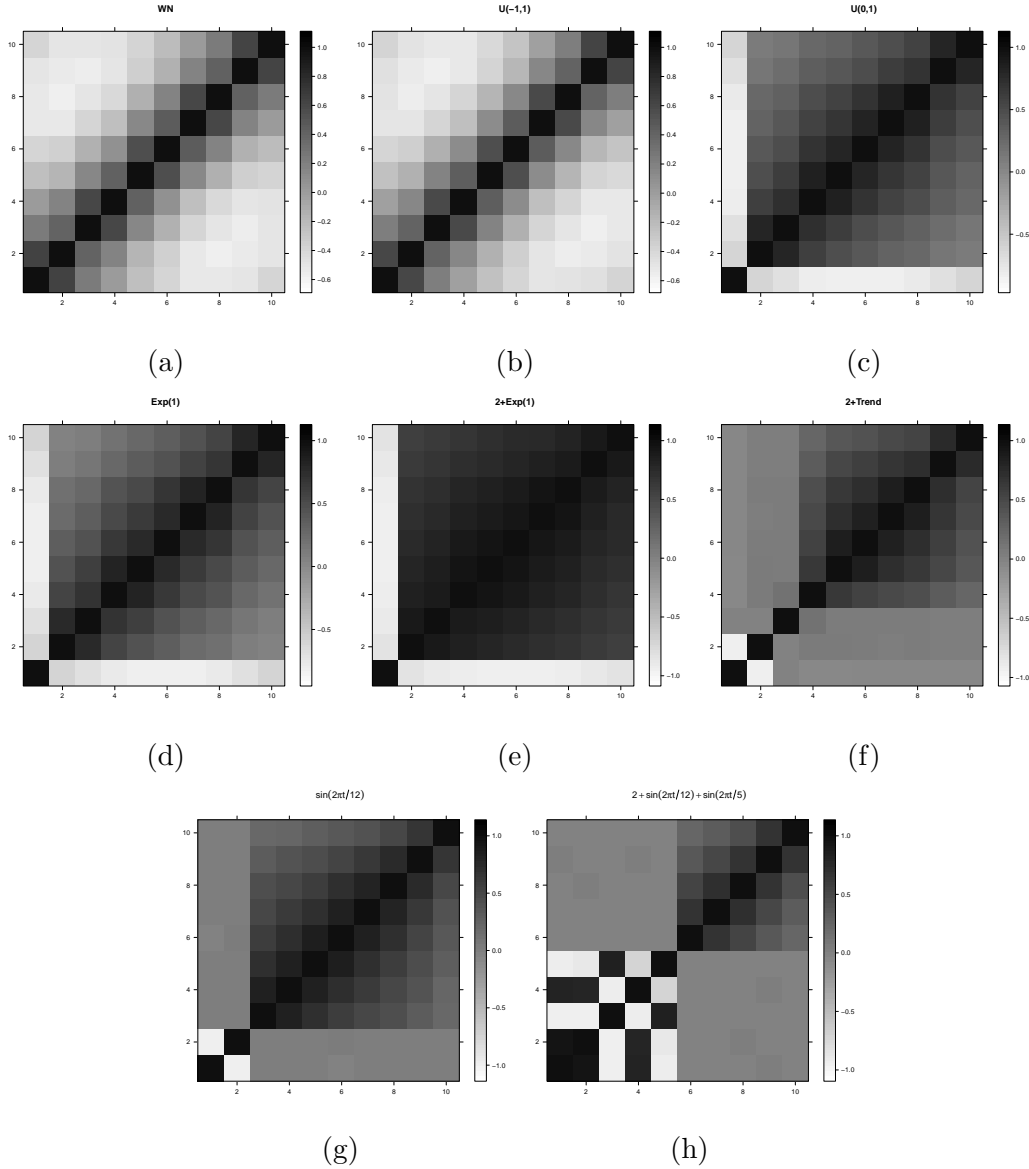


FIGURE 4.8: Matrix of correlation between ζ_i and ζ_j , $i, j = 1, \dots, 10$ for all cases.

The results clearly indicate that the elements of the main diagonal of the correlation matrix are more correlated than those on the off-diagonal for WN (see Figure. 4.8 (a)). Moreover, it can be seen that the value of the correlation becomes smaller

as the distance increases. Figure. 4.8 (c) illustrates the relationship between the ζ_i and ζ_j for uniform distribution (c). It is obvious that the relationship is totally different to that observed for the white noise series. Similar results were obtained for exponential case (b) and uniform distribution (d), respectively.

It is also clear that the values of ρ between ζ_1 and ζ_i , ($i = 2, 3, 4$) are positive for the white noise series and sample generated from uniform distribution (b). In contrast, there is a negative correlation between ζ_1 and ζ_i , ($i = 5, \dots, 10$) for these cases. Moreover, it is obvious that the strongest relationship is observed between ζ_1 and ζ_2 for the white noise process. It is also obvious that there is a negative relationship between ζ_1 and other ζ_i for cases (c) and (d). This further supports the results obtained above.

Furthermore, the correlation matrices for both exponential distributions, cases (d) and (e), are similar. This means adding a constant to exponential distribution (d) does not make differences to the results of the association. We observe that the correlation between ζ_1 and ζ_2 for sine wave (g) is approximately -1, which indicates a strong negative relationship. Similar results were achieved for the trend series. It can be observed that for sine series (h) there is a strong positive relationship between ζ_1 , and ζ_i ($i = 2, 4$), whilst ζ_1 is associated with ζ_3 and ζ_5 negatively. Furthermore, it can be seen that the values of ρ between ζ_i and ζ_j ($i, j = 6, \dots, 10$) are similar (see Figure. 4.8 (h)).

Figure. 4.9 shows the correlation matrix between ζ_i and ζ_j for WN considering different types of correlations, namely Pearson (ρ), Spearman (ρ_s) and Kendall (τ) (see Appendix A for details). It is clear that these measures give similar results and patterns of the correlation matrix. The results indicate that we can use any linear measure as there is no difference in the results. The results also clearly indicate that the elements of the main diagonal of the correlation matrix are more correlated than those of the off-diagonal. The values of the correlation obtained by the different measures become smaller as the distances increase.

Additionally, we also consider a measure based on mutual information that can be helpful in capturing serial dependence and correlation when there are nonlinearities

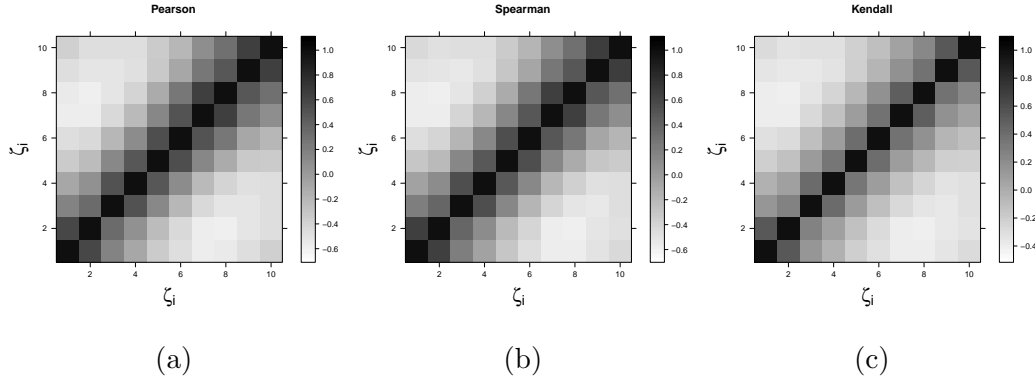


FIGURE 4.9: Matrix of correlation between ζ_i and ζ_j , $i, j = 1, \dots, 10$ for the WN process considering Pearson, Kendal, and Spearman correlations.

in the data. The standard measure for the mutual information (ξ) is defined in Appendix A, Eq. (A.6), which can capture the overall dependence, both linear and nonlinear, between ζ_i and ζ_j . Note that it takes values between 0 and 1. The results indicate that the relationship between ζ_i and ζ_j observed by the standard mutual information ξ , is similar to what was obtained by the linear correlation coefficients. For example, Table. 4.3 represents the standard mutual information between ζ_1 and other ζ_i for all cases. Accordingly, as the value of linear correlation informs us the positive and negative direction of association, we use linear correlation in the following chapters.

4.6 The effect of the mean and variance of Gaussian noise distribution

In the previous sections the white noise series ($\mu = 0$ and $\sigma^2 = 1$) and other cases were assessed. Let us now consider the effect of the Gaussian noise mean, μ , and variance, σ^2 , on the distribution of ζ_i . The results are provided only for ζ_1 , ζ_5 and ζ_{10} as similar results are observed for other cases. Let us first consider the situation where $\mu = 0$ and σ^2 varies. Figure. 4.10 shows the results related to $\bar{\zeta}_i$, σ_i^2 , and CV_i of ζ_i for the sample obtained from $N(0, 5)$. These results are similar to what emerged for the white noise process, $N(0, 1)$. The simulation results for other values of σ^2 also confirm that changing the values of σ^2 does not show any different

WN									
	ζ_2	ζ_3	ζ_4	ζ_5	ζ_6	ζ_7	ζ_8	ζ_9	ζ_{10}
ζ_1	0.92	0.88	0.79	0.75	0.80	0.84	0.87	0.90	0.85
$U(0, 1)$									
	ζ_2	ζ_3	ζ_4	ζ_5	ζ_6	ζ_7	ζ_8	ζ_9	ζ_{10}
ζ_1	0.91	0.86	0.82	0.76	0.83	0.84	0.88	0.89	0.88
$U(-1, 1)$									
	ζ_2	ζ_3	ζ_4	ζ_5	ζ_6	ζ_7	ζ_8	ζ_9	ζ_{10}
ζ_1	0.93	0.91	0.83	0.86	0.84	0.85	0.89	0.92	0.87
$Exp(1)$									
	ζ_2	ζ_3	ζ_4	ζ_5	ζ_6	ζ_7	ζ_8	ζ_9	ζ_{10}
ζ_1	0.87	0.89	0.91	0.92	0.94	0.93	0.89	0.88	0.86
$2 + Exp(1)$									
	ζ_2	ζ_3	ζ_4	ζ_5	ζ_6	ζ_7	ζ_8	ζ_9	ζ_{10}
ζ_1	0.89	0.93	0.94	0.96	0.97	0.98	0.95	0.93	0.91
$2 + Trend$									
	ζ_2	ζ_3	ζ_4	ζ_5	ζ_6	ζ_7	ζ_8	ζ_9	ζ_{10}
ζ_1	0.99	0.81	0.65	0.63	0.65	0.64	0.66	0.67	0.66
$sin(2\pi t/12)$									
	ζ_2	ζ_3	ζ_4	ζ_5	ζ_6	ζ_7	ζ_8	ζ_9	ζ_{10}
ζ_1	0.99	0.67	0.69	0.70	0.68	0.68	0.69	0.66	0.68
$2 + sin(2\pi t/12) + sin(2\pi t/5)$									
	ζ_2	ζ_3	ζ_4	ζ_5	ζ_6	ζ_7	ζ_8	ζ_9	ζ_{10}
ζ_1	0.99	0.96	0.98	0.95	0.58	0.57	0.56	0.57	0.57

TABLE 4.3: Non-linear correlation between ζ_1 and ζ_i for all cases.

results. Therefore, for $\mu = 0$ and $\sigma^2 \geq 1$, the patterns of ζ_i are similar to those which emerged for the white noise process $N(0, 1)$.

Let us now consider $\sigma^2 = 1$ (fixed) and different values of μ . The outcome indicates that if the sample is taken from $N(\mu \neq 0, 1)$, the distribution of ζ_i might be symmetric (see Figure. 4.11). In Figure. 4.11, the distribution of ζ_1 is symmetric based on the D-P test for the symmetry assumption. Similar results have been obtained for ζ_i , $4 \leq i \leq 9$. However, the distribution of ζ_{10} is skewed to the left, based on the D-P test (see also Figure. 4.11). Therefore, for the sample taken from $N(1, 1)$, for instance, the distributions of ζ_i are symmetric except ζ_2 , ζ_3 and ζ_{10} whose distributions are skewed. Similar results have been observed for the sample taken from $N(\mu \neq 0, 1)$. As a result, μ and σ^2 play an important role in identifying the symmetry of the distribution of ζ_i . Note also that one may use the standardisation procedure to transfer a series with arbitrary mean μ and variance σ^2 to a series with zero mean and also variance 1.

Additionally, a commonly used approach to determine if a variable follows a normal distribution is the normal Q-Q plot. Visually, one may decide whether the data of ζ_i roughly follows a straight line (see Figures 4.10 and 4.11 (bottom)). This may help us to identify which of them are normally distributed, and then which are symmetric. However, we use one of the most powerful test for all types of distributions and sample sizes, the Shapiro-Wilk test (S-W) [221], to evaluate the assumption of the normality of ζ_i . It confirms that if the distribution of ζ_i is normal, then it must be symmetric. Here, the simulation for the p -value of the S-W test, is also considered to have a good and credible result. The simulation was repeated 10^3 times. Table 4.4 reports the average of the p -value of the S-W test for ζ_i . The results indicate that the average of the p -value of ζ_5 and ζ_6 for the sample taken from $N(0, \sigma^2)$ is greater than 0.05. Thus, the distributions of ζ_5 and ζ_6 follow a normal distribution. However, the distributions of other ζ_i are not normally distributed; their p -values of the S-W test are very small, leading to a rejection of the normality assumption. Therefore, the distributions of ζ_i are not symmetric for the white noise process, except the distribution of ζ_5 and ζ_6 . For the sample taken from $N(\mu \neq 0, 1)$, the distributions of ζ_i might be symmetric based on the p -values of the S-W test. For example, if we select $\alpha = 0.05$, then ζ_i ($i = 1, 4, \dots, 9$) is normally distributed for sample $N(10, 1)$ as the p -value > 0.05 . It is noticed that if the sample is selected from $N(\mu, \sigma^2)$, where $\mu \neq 0$, the distribution of ζ_i can be symmetric.

Figure 4.12 displays the distribution of ζ_1 for different values of $\mu \neq 0$ and σ^2 . The histogram of ζ_1 for the sample generated from $N(1, 1)$ is symmetric, whilst it is skewed for the samples generated from $N(1, 50)$ and $N(100, 1)$.

The other empirical results confirm that there is a relationship between ζ_1 and ζ_2 whereas a very weak relation is seen between ζ_1 and other ζ_i for the sample taken from $N(0, 5)$. Similar results were obtained for the sample taken from $N(0, 10)$. Thus, the correlation between ζ_1 and ζ_i for $N(0, \sigma^2)$ is similar to that for the white noise process $N(0, 1)$. However, for the sample taken from $N(1, 1)$, there is a clear negative correlation between ζ_1 and other ζ_i .

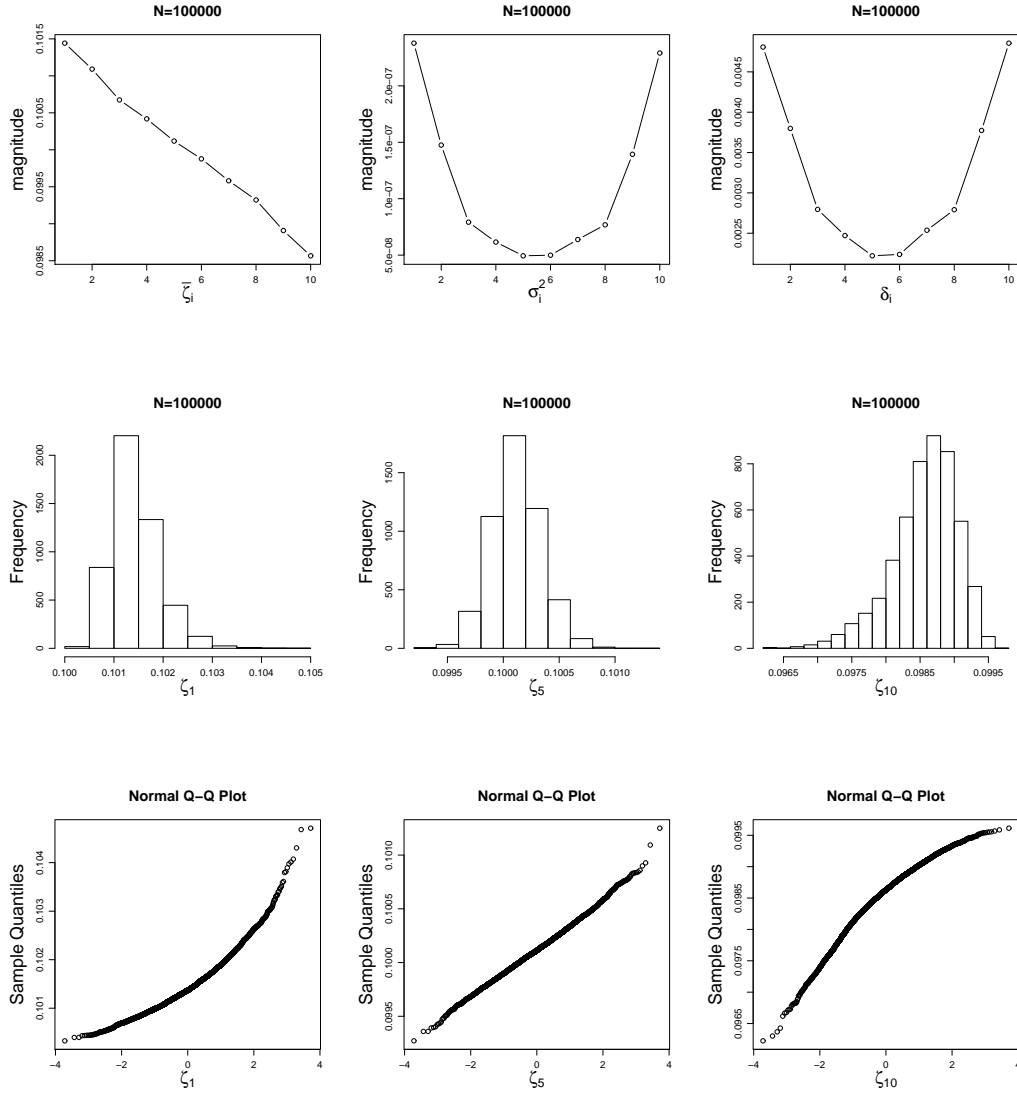


FIGURE 4.10: Plot of $\bar{\zeta}_i$, σ_i^2 and CV_i (top), the histograms of ζ_1 , ζ_5 , and ζ_{10} (middle) and the $Q-Q$ plots (bottom) for $\mu = 0$ and $\sigma^2 = 5$.

4.7 Effect of L on the distribution of the largest and smallest eigenvalues for WN

In this section, we consider the effect of L on the distribution of the largest and smallest eigenvalues of $\mathbf{B}_{L,N}$ for Y_N as a white noise. We consider different values of L to explore how changing the values of the window length affect the behaviour or the distribution of $\zeta_1^{L,N}$ and $\zeta_L^{L,N}$ for WN.

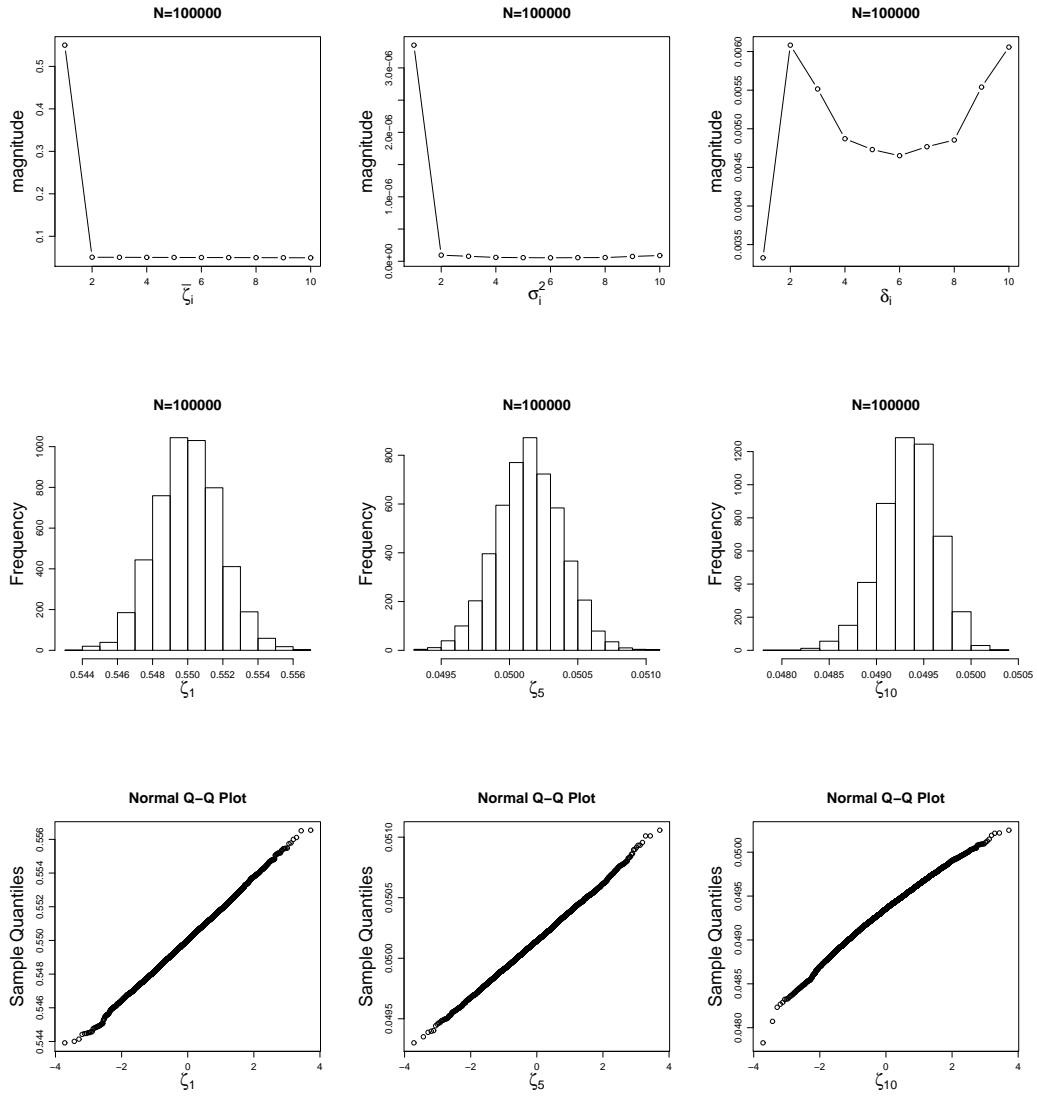


FIGURE 4.11: Plot of $\bar{\zeta}_i$, σ_i^2 and CV_i (top), histograms of ζ_1 , ζ_5 , and ζ_{10} (middle) and $Q-Q$ plots (bottom) for $\mu = 1$ and $\sigma^2 = 1$.

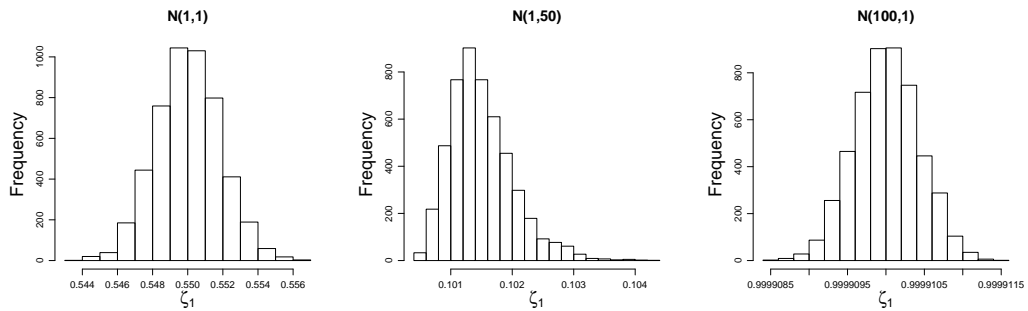


FIGURE 4.12: Histogram of ζ_1 for different values of μ and σ^2 .

	<i>p</i> -value of S-W test					
	$\mu = 0, \sigma^2 = 5$	$\mu = 0, \sigma^2 = 10$	$\mu = 0, \sigma^2 = 100$	$\mu = 1, \sigma^2 = 1$	$\mu = 5, \sigma^2 = 1$	$\mu = 10, \sigma^2 = 1$
ζ_1	1.9e-16	2.5e-18	1.5e-18	0.182*	0.011	0.360*
ζ_2	5.1e-11	3.1e-9	2.7e-10	2e-4	6.2e-4	6e-4
ζ_3	2.2e-4	1.3e-4	2.8e-5	9.4e-3	0.020	0.019
ζ_4	0.019	0.016	0.015	0.180*	0.162*	0.151*
ζ_5	0.055*	0.391*	0.252*	0.632*	0.184*	0.413*
ζ_6	0.110*	0.107*	0.104*	0.412*	0.324*	0.312*
ζ_7	0.005	0.004	0.003	0.450*	0.392*	0.381*
ζ_8	8.1e-4	1.3e-4	8.3e-8	0.514*	0.442*	0.451*
ζ_9	2.4e-7	2.3e-6	2.8e-7	0.372*	0.480*	0.473*
ζ_{10}	2.2e-16	2.2e-16	2.2e-16	0.0420	0.008	2.1e-05

Note :*, represents normality based on Shapiro test at $p = 0.05$.

TABLE 4.4: Average of p -values of the Shapiro-Wilk test of ζ_i for different values of μ and σ^2 of Gaussian distribution.

Let $Y_N = (y_1, \dots, y_N)$, where $y_t \sim \text{WN}(0, 1)$ and $N = 100$. We consider five different values of L , and for each value of L , the simulation is repeated 10^5 times to observe the behaviour or the distribution of the largest and smallest eigenvalues in each case. We intend to see if there is an effect on the eigenvalues by changing the window length; specifically, to see the similarities and differences between the distribution of $\zeta_1^{L,100}$ and $\zeta_1^{L-m,100}$, and between the distribution of $\zeta_L^{L,100}$ and $\zeta_L^{L-m,100}$ where $L = 50$ and $m = (10, 20, 25, 30, 40)$. Figure. 4.13 shows the distribution of the largest eigenvalue for different values of L . It is clear that there is no change in the distribution of $\zeta_1^{L,100}$ with L . The distribution is skewed to the right for all cases. However, it can be seen that the distribution of $\zeta_L^{L,N}$ changes as L increases (see Figure. 4.14). For $L = 10$, the distribution of $\zeta_L^{10,100}$ is skewed to the left, then becomes symmetric for $L = 30$, and becomes skewed to the right for $L = 50$.

The results confirm that the largest eigenvalue of \mathbf{B} has a stable positive skewed distribution with L whilst the smallest eigenvalue does not have a stable distribution with L ; the distribution changes from left to right when L increases. This result is very interesting and helpful for the following chapters, especially Chapter 6 because the results indicate that if we add any signal component to the noise series, for example harmonic, trend and cyclical components, then the first few eigenvalues are related to those components and as soon as we obtain the noise level the pattern of eigenvalues will be similar to those found for the noise series. Therefore, the results

obtained here are very important for time series and signal processing techniques where noise reduction, signal extraction and filtering matters.

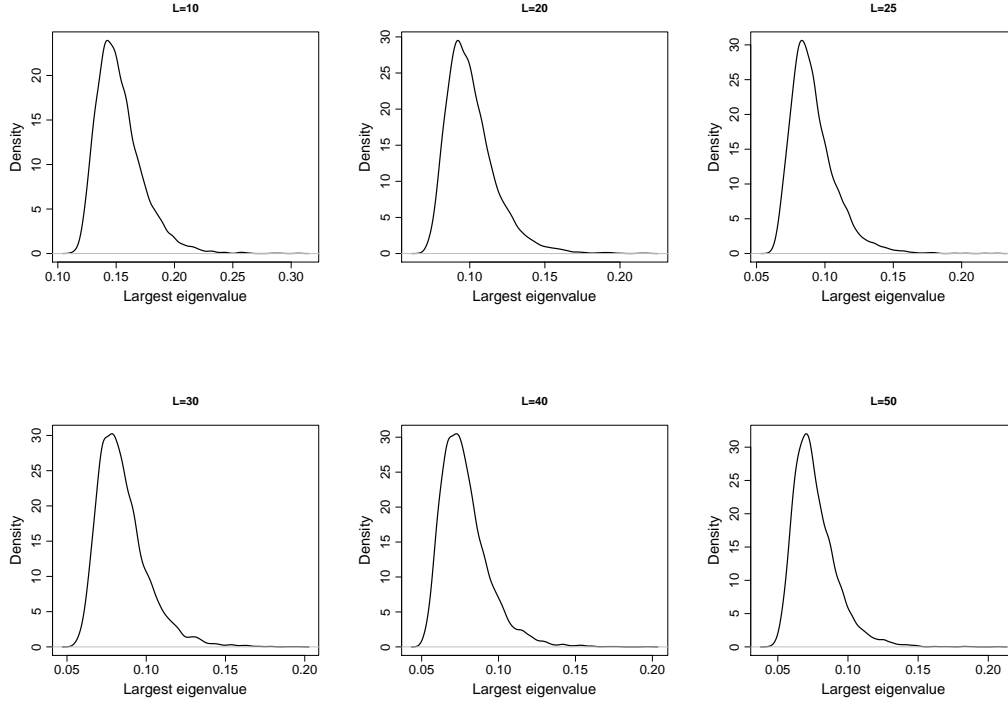


FIGURE 4.13: Plot of the largest eigenvalue for the white noise process with respect to different values of L .

4.8 Summary

In this chapter, we have studied the empirical distribution of ζ_i of the matrix \mathbf{B} , generated from different distributions (symmetric, nonsymmetric), trend and sine series. The results have illustrated that for a large sample size N , $\bar{\zeta}_i \rightarrow 1/L$, and $\sigma_1^2 \simeq \sigma_L^2$ for the normal and the uniform distributions with zero mean, whilst ζ_i does not converge to $1/L$, and $\sigma_1^2 \neq \sigma_L^2$ for other cases.

It was observed that for the symmetric cases, the first eigenvalue has a right skewed distribution, whilst it may have a symmetric distribution for the trend and nonsymmetric distribution examples. Moreover, for all cases considered in this chapter, the distribution of the middle ζ_i , for $L = 10$, can be symmetric except the distribution

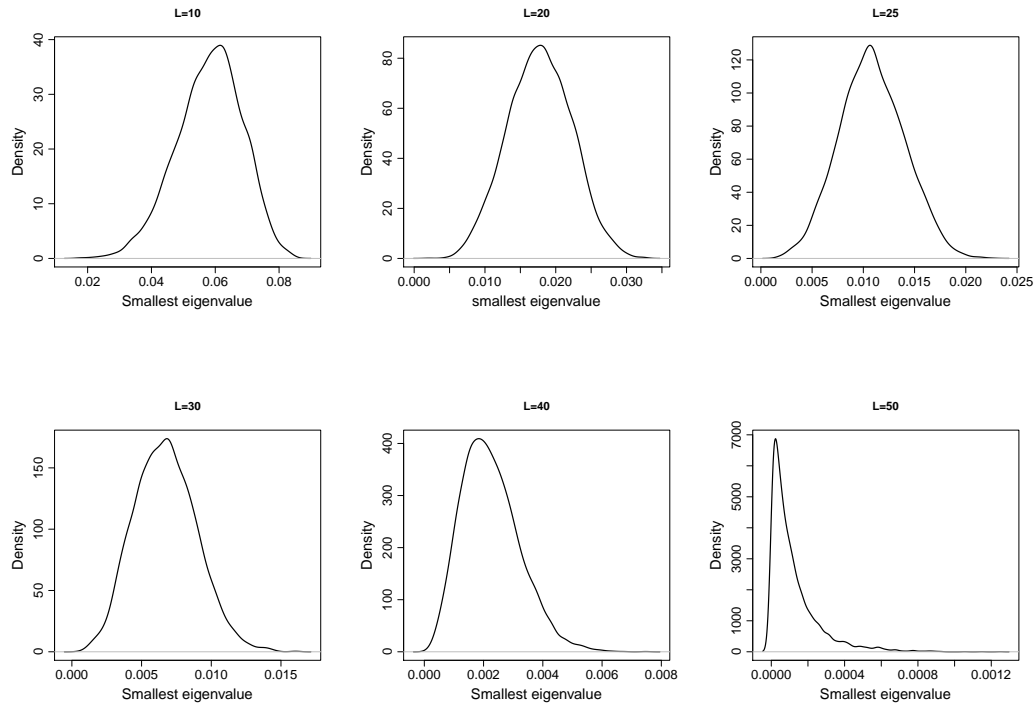


FIGURE 4.14: Plot of the smallest eigenvalue for the white noise process with respect to different values of L .

of ζ_5 for the trend case and sine series (g). It was found that the smallest eigenvalue has a negative skewed distribution for all cases apart from the trend series and sine waves. In addition, the correlation between eigenvalues was assessed and found that there is a complex structure among them. It is also found that the correlation matrix between ζ_i and ζ_j for symmetric distribution is different to what observed for nonsymmetric distribution, trend and sine series. Furthermore, the influence of μ and σ^2 on ζ_i were examined. The results show that the distribution of ζ_i for the sample generated from $N(0, \sigma^2)$ is similar to the distribution of ζ_i for the sample taken from $N(0, 1)$. However, the distribution of ζ_i is totally different, and might be symmetric for the sample taken from $N(\mu, \sigma^2)$, $\mu \neq 0$. Thus, the rate σ^2/μ plays an important role in identifying if ζ_i has a symmetrical distribution. In addition, we consider the effect of the window length on the distribution of the largest and smallest eigenvalues for the white noise process. The results show that the distribution of the largest eigenvalue is skewed to the right for any value of L , but the distribution of the smallest eigenvalue changes from the left to the right with L .

The results obtained here can be used for signal processing and time series techniques in terms of filtering and noise reduction. The results confirm that if we simply add an intercept, then the pattern and distribution of ζ_i will be changed. Generally, if we add more non-stochastic components to the noise series, then the first leader eigenvalues correspond to those components, and once we reach the noise level, the pattern of eigenvalues will be similar to those found for the noise series. Consequently, this will aid in identification of the value of r and then of signal extraction using a step by step procedure in chapter [6](#).

Chapter 5

Chaos and Noise

5.1 Introduction

Chaos theory is a branch of mathematics that focuses on the dynamical systems whose behavior is highly sensitive to small changes in initial conditions. This means slight alterations can lead to strikingly considerable consequences and this popularly referred to as the butterfly effect [55]. Slight differences in initial conditions, for instance, to those due to rounding errors in numerical computation, give widely diverging results for such dynamical systems, yielding long-term forecast of their behavior hard to be obtain in general [222]. Note that this can be happened even though these systems are deterministic. This means that the future behavior of systems is totally determined by their initial conditions, with no involvement of random components. In different words, the deterministic nature of these systems does not make them predictable. Chaotic systems are predictable for a while and then become random [222].

Edward Lorenz summarized the theory as chaos is when the present determines the future, but the approximate present does not approximately determine the future. Chaotic behavior can be found in several natural systems, for example, weather and climate [55]. This behavior also exists spontaneously in other systems with artificial components, such as road traffic. It can be studied through analysis of a chaotic

mathematical model, or through analytical techniques such as recurrence plots and Poincar maps. The theory has several applications in different disciplines, including engineering, biology, meteorology, sociology, physics, environmental science, computer science, economics, ecology, and philosophy.

The systems behavior is known as chaos. Chaos term is known commonly as a state of disorder, but this term is defined more precisely in chaos theory. Despite there is no universal mathematical definition of chaos in the literature, the most used one says that to classify a dynamical system as chaotic, it must be sensitive to initial conditions, topologically mixing and have dense periodic orbits. For more information about the mathematical definition and formulation of chaos, see [223].

Some dynamical systems, for instance, Logistic map is chaotic everywhere, but in various cases chaotic behavior can be seen only in a subset of phase space. The most interesting cases arise when the chaotic behavior takes place on an attractor. The attractor that appear from chaotic systems is known as a strange attractor. Strange attractors occur in both continuous dynamical systems (such as the Lorenz system) and in some discrete systems (such as the Hnon map). The Lorenz attractor (see Figure 1.2 in Chapter 1) is probably one of the most and best known chaotic system diagrams, perhaps because it is one of the first and the most complex ones that shows a very interesting pattern, which looks like the wings of a butterfly.

In the late 19th century, while Henri Poincar was studying the three-body problem, he found that we can have orbits that non-periodic, and yet not forever increasing nor approaching a fixed point [224]. The start of chaos theory was in the field of ergodic theory, then on the topic of nonlinear differential equations were carried out, see, for example, [225] and references therein.

Although the first insights of chaos theory were in the first half of the 20th century, it became formalized only after the mid of the century, when scientists found at that time linear theory and the prevailing system theory could not describe the observed behavior of certain experiments such that of the logistic map. What was attributed to measure noise and imprecision were considered by chaos theorists as a full component of the studied systems.

The early pioneer of the chaos theory was Edward Lorenz. The main interest of his work in chaos came up accidentally in 1961 when he was working on weather forecast. He found that very small changes in initial conditions will produce large changes in long-term outcome [55]. His discovery, which is known as Lorenz attractors, displayed generally that even detailed atmospheric modeling does not make accurate long-term weather forecasts. In 1982 the fractal geometry of nature was published in [226], which became a classic of chaos theory. Biological systems such as the branching of the circulatory and bronchial systems proved to fit a fractal model [227].

In 1987, a book with title chaos: making a new science, was published by James Gleick [228], which became a best known one that introduce fundamental principles of the theory of chaos and its history to the broad public. Since then, it has more attention and progressively emerged as a discipline, mainly under the name of non-linear systems analysis. Currently, the theory remains an active area of research including several various disciplines [228]. Although it was born from observing weather patterns, its applicability has become known in a variety of areas. Some areas that benefit from chaos theory are geology, mathematics, microbiology, biology, computer science, economics, engineering, finance, algorithmic trading, meteorology, philosophy, physics, politics, population dynamics, psychology, and robotics. For example, chaotic behaviour is found in biological systems, such as cardiotocography, which is a technical means of recording the fetal heartbeat and the uterine contractions during pregnancy. Fetal surveillance is a delicate balance of obtaining precise information while being as noninvasive as possible. Chaotic modeling can help in obtaining better models of warning signs of fetal hypoxia [229].

In the last 3 decades, there have been several discussions on the similarities and differences between stochastic noises and deterministic chaos. The deterministic system, chaos, stochastic process and noise can exist in various situations [89]. Chaos is a nonlinear deterministic process which looks random. There have been great challenges for the detection of chaos since the seminal papers by [230, 231]. The distinction between chaos and noise systems is considered to be a very important topic, because they have several common properties that can make them

indistinguishable (see for example [90]–[96]). Furthermore, there is a significant and confusing similarity between these systems, in that both have an even distribution property in the time domain and a broad band property in the frequency domain [89].

In order to introduce this issue, let us present typical trajectories of two well known processes, chaotic and stochastic ones. Figure 5.1 (left) shows a series of nonlinear difference equation proposed by Henon for the parameter values $a = 1.4$ and $b = 0.3$, which appears to be a random series; however, mapping this series to a two dimensional map can present the chaotic attractor shown in Figure 5.2. In Figure 5.1 (right) a series is plotted from Gaussian white noise of zero mean and unit variance. Both series appear irregular and similar on time. However, there are important differences between them and, in particular; the Henon is known as deterministic and being more regular than the white noise. The visual comparison exposes that chaotic signal is similar to and more regular than the white noise. From this observation, one would wonder if there exists a quantitative test of randomness that can capture our visual intuition. Such a question is very important in the natural sciences where the time evaluation of several biological and physicochemical phenomena is described by stochastic process.

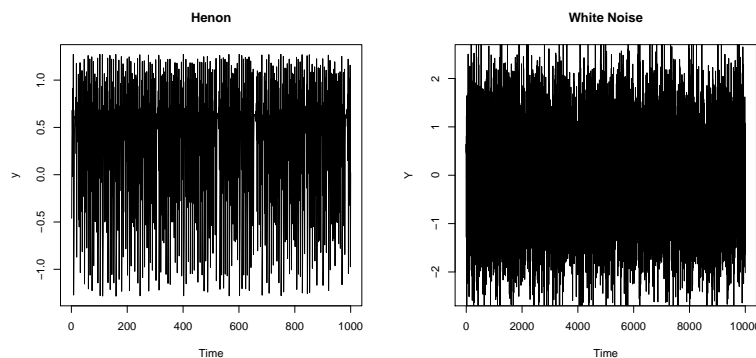


FIGURE 5.1: A Henon series (left) and white noise series (right).

As stated earlier, chaos theory has applications in several disciplines. Medicine is the latest field that consider the idea of chaos theory. Specialists are seeking to understand, for example, the working of the brain and heart using it, which may help them to understand and treat some medical conditions. An example of chaos

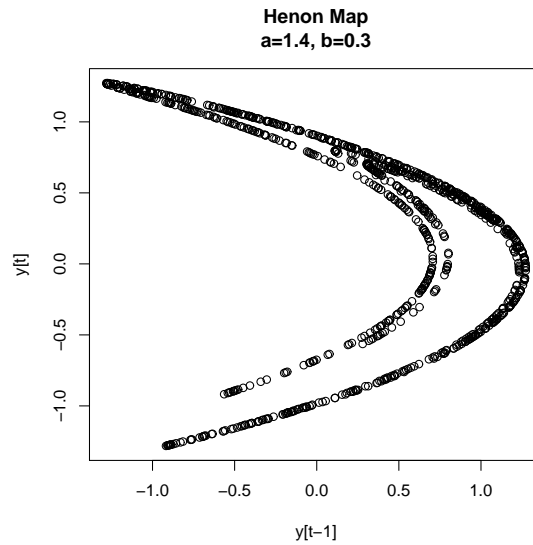


FIGURE 5.2: A Henon map.

in our bodies is the beating of the heart [232]. It is not constant, and depends on the actions that a person is doing at a certain of time. The heartbeats can speed up, or beat erratically, in certain different conditions.

The noise influencing the ECG time series has caused many problems in the estimation of chaotic behaviour and needs to be taken into account [233]. The issue of noise impact on the estimation of parameters for chaotic characterisation (e.g. dimensions, entropies and Lyapunov exponents) has been the subject of much work in recent years.

Whether the normal or abnormal human heart is chaotic has long been a subject of interest in the application of nonlinear time series analysis. For example, the author of [234, 235] analysed the ECG of healthy subjects and those with severe congestive heart failure. The results of their work suggest that cardiac chaos is prevalent in the normal heart, and a decrease in such chaotic behaviour can be a signal of congestive heart failure.

The correlation dimension, entropies and Lyapunov exponents (LE), show evidence of chaos in human cardiac data [233], [236]–[238]. For example, the authors of [236] combined ECG results and pulse data with periodic and chaotic data, and found that the changing tendency of ECG and pulse pressure signals are both consistent

with that of chaotic Rössler and Chua data and vary significantly from the considered periodic data, which indicates the existence of deterministic chaos in the cardiac data employed. The authors mentioned that their results may not constitute a definitive proof of chaos in healthy human heart output signals but they were found to be consistent with chaotic dynamics. Furthermore, the authors of [237] concluded their research by saying that the normal human heart follows a deterministic dynamic of chaotic nature.

The LE is the common key to characterise chaotic systems [92], with chaos indicated by a positive LE . However, it is not directly clear how it should be properly defined, and it suffers from some drawbacks [90]. Furthermore, evidence of chaos based on these measures, such as the presence of the positive Lyapunov exponent alone, do not provide adequate evidence to confirm the existence of chaotic behaviour [239]. The Pseudo-Periodic Surrogates (PPS) method and correlation dimension were used in [236], together with the hypotheses of a periodic orbit with uncorrelated noise in both complete ECG and pulse waveform, to examine the existence of chaotic structures. It was found that the ECG systems reveal chaotic characteristics. Here, the distribution of ζ_i is proposed to be a technique for distinguishing between these systems. The general aim is to analyse the experimental data and to detect whether their nature is purely random or due to deterministic chaos.

It should be noted that for over twenty years the issue of distinguishing random signals from deterministically chaotic ones has been debated in the nonlinear dynamics community (see for example, [240]-[242]). The question of whether the data are purely random or purely deterministic is not valid in most practical cases. In realistic situations, the observed processes originate from nonlinear systems, but are strongly affected by random noise. In such situations, the noise-free underlying dynamical system can demonstrate deterministic chaos, but being influenced by noise it can show a complex mixture of chaos and noise. The question is therefore not to decide whether the process is random or not (it is always random to some extent in practice), but whether the underlying noise-free dynamical system has deterministic chaos or not. An altogether different issue is whether random noise can induce chaos if the noise-free system behaves in a non-chaotic manner [243].

Based on the literature, the main issue is how we can determine whether the irregularity observed in the time series results from chaotic systems or from the noisy systems. Here, we apply the proposed method to distinguish between the white noise and some chaotic series, namely; the Henon, Logistic, Lorenz, and Tent map. The method is then examined using real ECG data.

In the following section, we give a general mathematical definition of chaos, and a brief definition of the Henon, Logistic, Lorenz and Tent map. In section 5.3, the distinction between chaotic and noise series is discussed using simulated and real chaotic series.

5.2 Mathematical definition of chaos

There are different definitions of chaos in the means of mathematics. Here we give one of the most popular definitions: for more information see [244].

A continuous map $f : X \rightarrow X$ on a compact metric space (X, d) is called chaotic according to Li and Yorke [245], if there exists a subset S of X with the following properties:

- $\limsup_{n \rightarrow \infty} d(f^n(x), f^n(y)) > 0$ for all $x, y \in S, x \neq y$.
- $\limsup_{n \rightarrow \infty} d(f^n(x), f^n(y)) = 0$ for all $x, y \in S, x \neq y$.
- $\limsup_{n \rightarrow \infty} d(f^n(x), f^n(p)) > 0$ for all $x \in S, p \in X, p$ periodic.

There are several chaotic maps; here we consider the most famous and well known ones.

1. The Henon map: a Henon map is a simple nonlinear difference equation:

$$x_{n+1} = y_n + 1 - ax_n^2 \quad (5.1)$$

$$y_{n+1} = bx_n. \quad (5.2)$$

Henon [246] used the parameters a and b as $a = 1.4$ and $b = 0.3$. Although those values are usually used, the other parameters are interesting too.

2. The Logistic map: in a mathematical sense, a logistic map is a nonlinear difference equation, given as follows:

$$x_{n+1} = rx_n - x_n. \quad (5.3)$$

where $x_n \in [0, 1]$, which represents the ratio of the existing population to the maximum possible population. The values of interest for the parameter r are those in the interval $(0, 4]$.

3. The Lorentz map: Lorentz [55] published his system of the three ordinary equations:

$$\frac{d(x)}{d(t)} = \sigma(y - x) \quad (5.4)$$

$$\frac{d(y)}{d(t)} = x(r - z) - y \quad (5.5)$$

$$\frac{d(z)}{d(t)} = xy - bz, \quad (5.6)$$

where σ , r , and b are positive parameters. His equations show the convective motion of a fluid cell which is warmed from below and cooled from above. He found a strange attractor when he used the values 10, 28 and $8/3$ for the parameters σ , r and b respectively.

4. The Tent map: a Tent map is one of the main examples of chaotic maps study for nonlinear discrete dynamical systems. It is given by the following formula:

$$x_{n+1} = f_a(x_n) = \begin{cases} ax_n & \text{for } x_n < \frac{1}{2} \\ a(1 - x_n) & \text{for } \frac{1}{2} \leq x_n, \end{cases}$$

where a is a positive real constant. Note that for $a = 4$, the Tent map is a nonlinear transformation of the Logistic map.

5.3 Distinction between chaos and noise

5.3.1 Simulated data

We simulated ζ_i ($i = 1, 2$) 1000 times for chaotic and white noise series. For the Lorenz series, the parameters of the system were $\sigma = 16$, $r = 45.92$ and $b = 4$, with an integral step of 0.01. The parameters $a = 1.4$ and $b = 0.3$ are considered for the Henon series, $r = 4$ for the Logistic, and $a = 2$ for the Tent map. The initial values were generated randomly from the uniform distribution to generate a chaotic series.

The empirical distribution of ζ_i , $i = 1, 2$ for the white noise, Logistic, Henon, Lorenz, and Tent series is evaluated. Note that to embed data from the Logistic equation or Henon map, embedding dimension $L = 2$ is sufficient. However, the embedding dimension for a chaotic Lorenz system must be at least 3. Moreover, according to the famous Takens theorem, the necessary embedding dimension should be calculated from the formula specified in [86]. For simplicity and visualisation purposes, $L = 2$ is considered. Here, we present five figures, with each figure showing four patterns for each case. The first is the plot of the time series data, the second displays the map, and the third and fourth patterns show the empirical distribution of ζ_i ($i = 1, 2$) for $m = 1000$ simulations. Looking at the plots of the series, particularly the white noise, Henon and Logistic series, it is difficult to decide whether the process is deterministic or stochastic. However, finding the distribution of ζ_i explained in the previous chapter, one can decide whether the data is chaotic or noise. Visually, we can judge that the distribution of ζ_i ($i = 1, 2$) is skewed for the white noise series (see Figure 5.3) and Tent series (see Figure 5.7). For the white noise series, the distribution of ζ_1 is skewed to the right and the distribution of ζ_2 is skewed to the left, whereas ζ_1 has a negative distribution and ζ_2 has a positive one for the Tent series. Note that ζ_i ($i = 1, 2$) might have a symmetric distribution for other chaotic series, especially the Henon and Logistic ones (see Figures. 5.4 and 5.5).

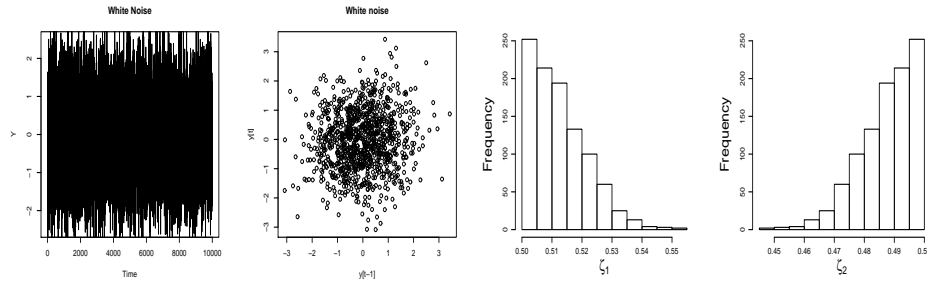
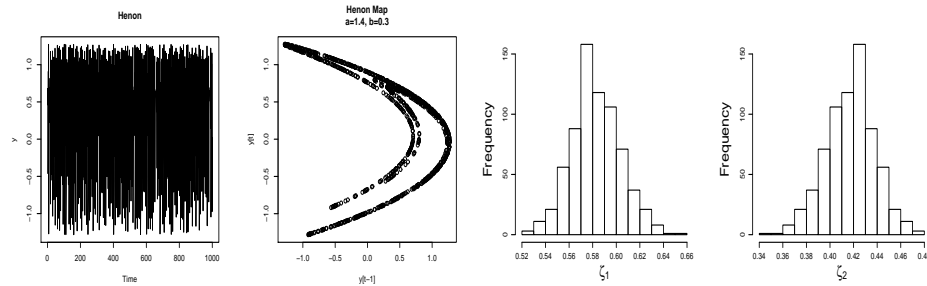
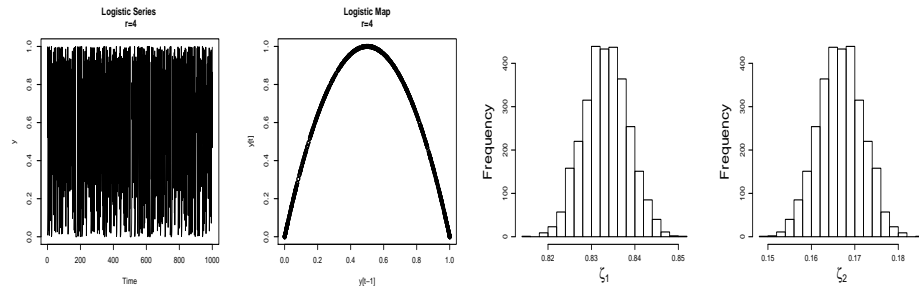
Furthermore, the coefficient of skewness, $skew(\zeta_i)$, is calculated. Table 5.1 presents the coefficient of skewness of ζ_i ($i = 1, 2$) for all cases. It is clear from the table that, for the white noise and Lorenz series, the coefficient of skewness for ζ_1 and ζ_2 are

approximately 1 and -1 respectively, whilst they are around zero for the Henon and Logistic series. Moreover, the D-P normality test is applied here to evaluate this problem. Table 5.2 provides information about the p -values of the D-P test for ζ_i . It is obvious that the p -value is not significant; greater than 0.01 for ζ_i , for the Henon and Logistic series, whilst less than 0.01 for the white noise, Tent and Lorenz series. Therefore, we accept the null hypothesis that the data of ζ_i , for the Henon and Logistic series are not skewed, and are therefore symmetric. For the Lorenz series, the empirical distribution might be similar for $L = 2$, but is completely different for $L = 3$ which is the correct dimension.

In addition to the previous results, the correlation ρ_s , between ζ_1 and ζ_2 is also computed. The value of ρ_s between ζ_1 and ζ_2 is exactly -1 for all cases, which indicates a perfect negative relationship between them. We also consider $L = 10$ for further evaluation. Figure 5.8 demonstrates the correlation matrix for ζ_i considering ρ_s . It is obvious that the correlation matrix between ζ_i and ζ_j , ($i, j = 1, \dots, 10$) for WN is totally different to what was observed for the chaotic time series. For example, there is a strong positive linear relationship between ζ_1 and ζ_2 for the WN series, whilst ζ_1 is associated with ζ_2 negatively for the chaotic series. Furthermore, for the WN series, the elements of the main diagonal of the correlation matrix are more correlated than those on the off-diagonal, whereas different structure among these elements can be seen for the chaotic series.

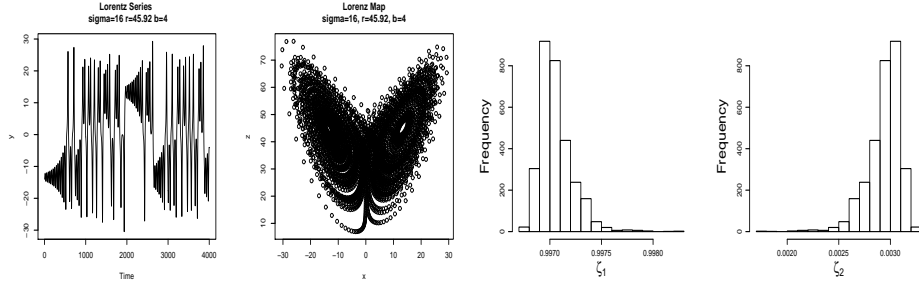
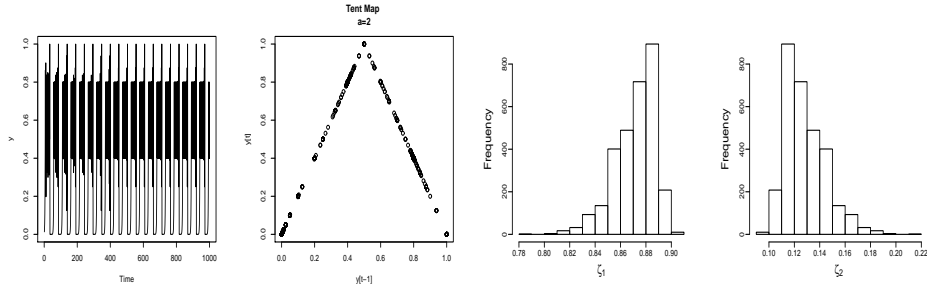
We also consider the nonlinear measure (ξ) between the eigenvalues, defined in Eq. (A.6). The ξ measure between ζ_1 and ζ_j ($j = 2, \dots, 10$), is evaluated. Table 5.3 presents the values of ξ for all cases. The linear and nonlinear correlations between ζ_1 and ζ_j , for all cases are similar. The results indicate that the relationship between ζ_1 and ζ_j observed by the standard mutual information ξ is similar to what was obtained by ρ_s . As a result, we should use the value of linear correlation because it gives us, the positive and negative direction of correlation.

Additionally, the Kolmogorov Smirnov (K-S) [247] test is also applied. The K-S test is a nonparametric test for the equality of continuous, one-dimensional probability distributions that can be used to compare a sample with a known probability

FIGURE 5.3: White Noise series, map and histogram of ζ_i ($i = 1, 2$).FIGURE 5.4: Henon series, Henon map and histogram of ζ_i ($i = 1, 2$).FIGURE 5.5: Logistic series, Logistic map and histogram of ζ_i ($i = 1, 2$).TABLE 5.1: Coefficient of skewness for ζ_i , ($i = 1, 2$) for white noise and chaotic series.

	Coefficient of Skewness of ζ_i .				
	WN	Henon	Logistic	Lorenz	Tent
ζ_1	0.99	-0.11	0.005	0.82	-0.93
ζ_2	-0.99	0.11	-0.005	-0.82	0.93

distribution, for instance normal or uniform distributions, or to compare two samples. The two-sample K-S test is one of the most helpful and general nonparametric methods for comparing two samples because it is critical to differences in both the location and shape of the empirical cumulative distribution functions (c.d.f.'s) of

FIGURE 5.6: Lorenz series, Lorenz map and histogram of ζ_i ($i = 1, 2$).FIGURE 5.7: Tent series, Tent map and histogram of ζ_i ($i = 1, 2$).TABLE 5.2: p -value of the D-P test for ζ_i , ($i = 1, 2$) for white noise and chaotic series.

	P-value of D-P test for ζ_i				
	WN	Henon	Logistic	Lorenz	Tent
ζ_1	<0.01	0.2	0.96	<0.01	<0.01
ζ_2	<0.01	0.2	0.96	<0.01	<0.01

the two samples. Suppose that x_1, \dots, x_{N_1} are observations on i.i.d random variables X_1, \dots, X_{N_1} with a c.d.f. $F_1(x) = \mathbb{P}(X \leq x)$ and y_1, \dots, y_{N_2} are observation on i.i.d. random variables Y_1, \dots, Y_{N_2} with a c.d.f. $F_2(y) = \mathbb{P}(Y \leq y)$. The aim is to test

$$H_0 : F_1 = F_2 \text{ vs. } H_1 : F_1 \neq F_2.$$

Let $F_{1,N_1}(x)$ and $F_{2,N_2}(y)$ be the corresponding empirical c.d.f's;

$$F_{1,N_1}(x) = \mathbb{P}_{N_1}(X \leq x) = \frac{1}{N_1} \sum_{i=1}^{N_1} I(X_i \leq x)$$

$$F_{2,N_2}(y) = \mathbb{P}_{N_2}(Y \leq y) = \frac{1}{N_2} \sum_{i=1}^{N_2} I(Y_i \leq y).$$

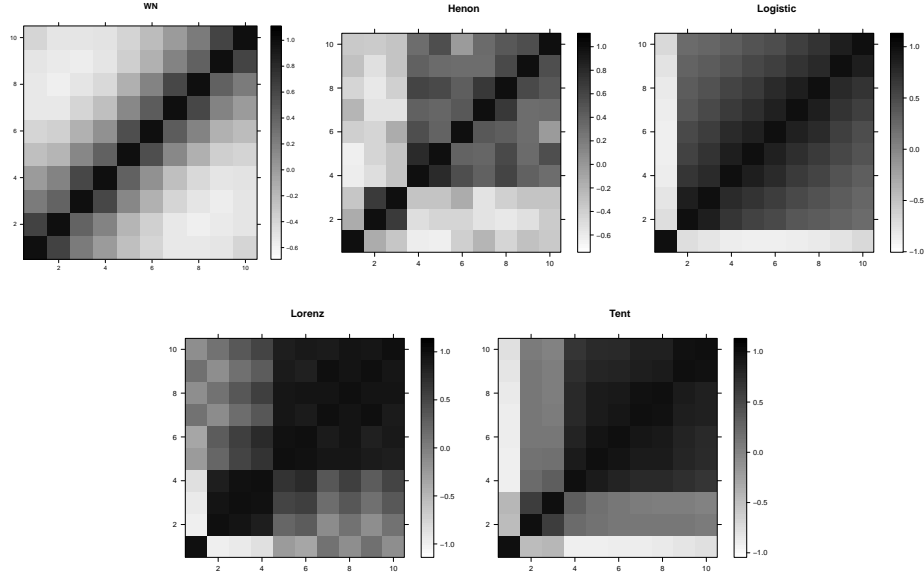


FIGURE 5.8: Matrix of correlation between ζ_i and ζ_j , $i, j = 1, \dots, 10$ for the white noise and chaotic series.

TABLE 5.3: Mutual information $I(\zeta_1, \zeta_i)$ ($i = 2, \dots, 10$) and the standard measure for mutual information $\xi(\zeta_1, \zeta_i)$ for white noise and chaotic series.

ζ_1 WN									
	ζ_2	ζ_3	ζ_4	ζ_5	ζ_6	ζ_7	ζ_8	ζ_9	ζ_{10}
ξ	0.94	0.85	0.83	0.84	0.85	0.86	0.86	0.86	0.85
I	1.02	0.63	0.58	0.59	0.59	0.62	0.68	0.67	0.64
ζ_1 Henon									
	ζ_2	ζ_3	ζ_4	ζ_5	ζ_6	ζ_7	ζ_8	ζ_9	ζ_{10}
ξ	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.98
I	0.51	0.60	0.72	0.73	0.59	0.52	0.61	0.56	0.64
ζ_1 Logistic									
	ζ_2	ζ_3	ζ_4	ζ_5	ζ_6	ζ_7	ζ_8	ζ_9	ζ_{10}
ξ	0.99	0.90	0.90	0.93	0.94	0.93	0.92	0.91	0.88
I	0.79	0.87	0.99	1.04	1.07	1.03	0.96	0.82	0.77
ζ_1 Lorenz									
	ζ_2	ζ_3	ζ_4	ζ_5	ζ_6	ζ_7	ζ_8	ζ_9	ζ_{10}
ξ	0.99	0.98	0.86	0.86	0.85	0.85	0.86	0.86	0.84
I	2.87	1.46	1.23	0.65	0.72	0.65	0.72	0.68	0.66
ζ_1 Tent									
	ζ_2	ζ_3	ζ_4	ζ_5	ζ_6	ζ_7	ζ_8	ζ_9	ζ_{10}
ξ	0.87	0.89	0.94	0.96	0.95	0.95	0.94	0.93	0.92
I	0.67	0.73	1.03	1.22	1.15	1.15	1.12	0.97	0.95

Under the null hypothesis, the largest difference between F_1 and F_2 goes to 0 in probability:

$$D_{N_1, N_2} = \sup_{x \in \mathbb{R}} |F_{N_1}(x) - F_{N_2}(x)| \rightarrow 0.$$

The idea in the K-S test is that the distribution of the supremum does not depend on the unknown distribution \mathbb{P} of the sample if it is continuous distribution.

Theorem 5.1. *Let X_1, \dots, X_{N_1} and Y_1, \dots, Y_{N_2} be i.i.d random variables with a common continuous c.d.f and let F_{1, N_1} and F_{2, N_2} be the empirical c.d.f's of X 's and Y 's, respectively. Moreover, let*

$$D_{N_1, N_2} = \sup_x |F_{N_1}(x) - F_{N_2}(x)|.$$

Then we have

$$\lim_{N_1, N_2 \rightarrow \infty} \mathbb{P} \left(\sqrt{\frac{N_1 N_2}{N_1 + N_2}} \leq x \right) = 1 - 2 \sum_{s=1}^{\infty} (-1)^{s-1} e^{-2s^2 x^2}.$$

In our case, the K-S statistic quantifies a distance between the empirical distribution functions of two samples, with ζ_i obtained from the white noise process and ζ_i obtained from the chaotic series. The null distribution of this statistic is computed under the null hypothesis that the samples are drawn from the same distribution. Table 5.4 shows the p -value of the K-S test between ζ_i for WN and ζ_i for each chaotic series. It is obvious that the null hypothesis is rejected since the p -value is less than 0.05, which is a default value of the level of significance for all cases. According to the test, the difference between ζ_i for WN and ζ_i for chaotic series is highly significant and we can say that they have a different distribution.

TABLE 5.4: p -value of the K-S test between ζ_i for the white noise series and ζ_i for chaotic series.

The p -value of K-S test				
	ζ_1 (Henon)	ζ_1 (Logistic)	ζ_1 (Lorenz)	ζ_1 (Tent)
ζ_1 (WN)	$< 2.2 \times 10^{-16}$	$< 2.2 \times 10^{-16}$	$< 2.2 \times 10^{-16}$	$< 2.2 \times 10^{-16}$
	ζ_2 (Henon)	ζ_2 (Logistic)	ζ_2 (Lorenz)	ζ_2 (Tent)
ζ_2 (WN)	$< 2.2 \times 10^{-16}$	$< 2.2 \times 10^{-16}$	$< 2.2 \times 10^{-16}$	$< 2.2 \times 10^{-16}$

5.3.2 Real data

To apply the proposed technique to real data one needs to generate several copies of the data sets. To this end, we utilise a simulation approach. By performing simulations and analysing the generated series, we can gain an understanding of how the pattern of eigenvalues would behave or what the histogram of each eigenvalue would be. Similarities and dissimilarities with the chaotic series provide an insight into detecting chaotic behaviour. Let us now consider an electrocardiogram (ECG) data, which are a record of the electrical activity of the heart.

ECG is a time series of the electrical potential between two points on the surface of the body caused by a beating heart. This time series or the signal is recorded by a machine to see if there is unusual event or problem in the heart signal. It is arguably one of the the most important time series. ECG is a simple test that can be used to check the heart's rhythm and electrical activity. Sensors attached to the skin are used to detect the electrical signals produced by the heart each time it beats. The ECG data can help to investigate symptoms of a possible heart problem, such as chest pain, suddenly noticeable heartbeats (palpitations), dizziness and shortness of breath. It can also help to detect arrhythmias; where the heart beats too slowly, too quickly, or irregularly, coronary heart disease; where the heart's blood supply is blocked or interrupted by a build-up of fatty substances, heart attacks; where the supply of blood to the heart is suddenly blocked, and cardiomyopathy; where the heart walls become thickened or enlarged.

The data of ECG signal analysed here was obtained via [248] and analysed by many authors, see, for example, [249]. The signal is a complicated one as it contains anomaly and considered as a time series discord. It was annotated by a cardiologist as containing one premature ventricular contraction, see Figure 5.9 (left). Table 5.5 represents a summary of descriptive statistics for the ECG series. The mean and standard deviation (S.D) of the series are -0.176 and 0.838, respectively. As it can be seen from the table that the values of skewness and kurtosis are negative and positive respectively, which clearly are different from 0, and typically indicate that the distribution is asymmetrical, more peaked and fatter tails than normal

distribution. Similar results can also be seen for the values of the minimum and maximum of the series, respectively.

Figures 5.9 and 5.10 show the results of the ECG series. It is very clear that the scatter plot of y_t with time delay 22 has a tractive pattern. A tractive pattern can also be seen with time delay 230. The distributions of ζ_i ($i = 1, 2$) are symmetric based on the D-P test; ζ_1 is slightly skewed to the left whilst ζ_2 is to the right (see also Table 5.6). The p -value of the K-S test between ζ_i for the WN and ζ_i for the ECG series is significant (almost zero). This means the data of ζ_i for WN and ECG do not come from the same distribution. The results confirm that the distribution of ζ_i ($i = 1, 2$) for ECG are totally different from what emerged for the white noise process, and similar to the distribution of ζ_i for the Henon and Logistic series.

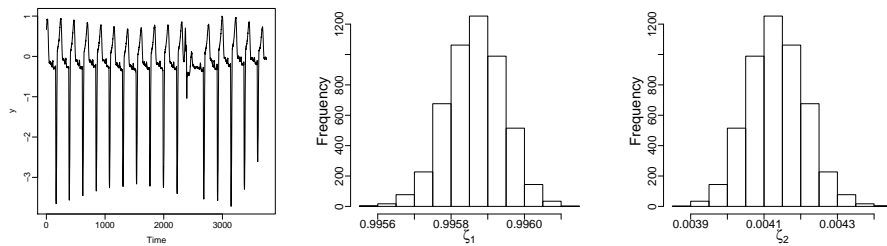


FIGURE 5.9: ECG series and histogram of ζ_i ($i = 1, 2$).

TABLE 5.5: Descriptive statistics of ECG series.

	Statistics					
	Mean	S.D	Minimum	Maximum	Skewness	Kurtosis
ECG series	-0.18	0.84	-3.72	1.00	-2.29	8.91

TABLE 5.6: Coefficient of skewness and p -value of the D-P test for ζ_i ($i = 1, 2$), for the ECG series.

	$Skew(\zeta_i)$	p -value of D-P test
ζ_1	-0.12	0.06
ζ_2	0.12	0.06

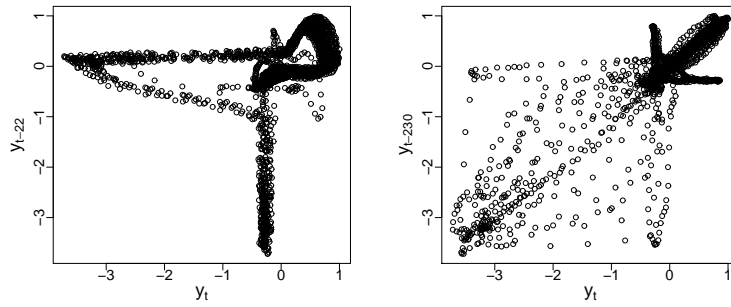


FIGURE 5.10: Plot of y_t with time delay 22 (left) and time delay 230 (right) for the ECG series.

5.4 Summary

The distribution of the eigenvalues of matrix \mathbf{B} generated from deterministic and stochastic processes were studied. The distribution of ζ_i was proposed to be a new approach to distinguishing the Henon, Logistic, Lorenz, and Tent chaotic series from the white noise series. It was found that the distribution of ζ_i is symmetric for the Henon and Logistic series, whilst skewed for the white noise, Lorenz, and Tent series. The distribution of ζ_i for white noise is totally different from what was obtained for the Tent map, as the direction of the skewness is different. Furthermore, the correlation between eigenvalues was assessed. It was observed that the correlation matrix between eigenvalues for the white noise series is totally different to what was emerged for the chaotic series. Although the distribution of ζ_i is skewed for the Lorenz series, it is possible to distinguish whether the process is deterministic or stochastic by calculating the coefficient of skewness for ζ_i or by looking at the matrix of the correlation. Furthermore, we applied the K-S test to compare the distribution of ζ_i for WN and chaotic series. The results indicate that the distribution of ζ_i for these have different distributions. The results of the real ECG time series also indicated that the proposed approach can be a valuable aid in distinguishing between chaos and noise.

Chapter 6

Selecting the Number of Eigenvalues

6.1 Introduction

In the previous chapters, we have studied and analysed the distributions and behaviour of the scaled Hankel matrix eigenvalues, and proposed these distributions for the identification of chaos from noise. The results make a significant contribution to achieving the goal of this chapter, which is the optimal choice of r and the extraction of the desired signal. In this regard, the purpose of this chapter is to develop the proposed approach and to indicate how it can be used directly to identify the eigenvalues corresponding to the noise component and thereby, selecting the appropriate value of r corresponding to the signal component.

Signal and noise separation, and reconstructed series are the main steps in SSA that need to be undertaken before any further application or aims; for example, forecasting, analysis of missing data, and change point detection. The main component (signal) can consist of subcomponents such as trend, and cyclical and seasonal ones. In the proposed approach, our interest is to consider the signal as a whole, so we seek to identify the optimal eigenvalues related to the whole signal component. In other words, we are not interested in each signal component, so the selection of the

window length rational to the periodicity of the signal components is not important [6]. In this case, we seek to illustrate how the optimal value of r can be achieved.

The structure of the signal or the time series is closely related to the rank of \mathbf{H} ($\text{rank } \mathbf{H}$), the number of non-zero eigenvalues in the SVD of the trajectory matrix \mathbf{H} . For any time series of finite length, $\text{rank } \mathbf{H} \leq \min(L, K)$. If $\text{rank } \mathbf{H} < \min(L, K)$, then the series has a structure [6]. This chapter considers several examples of time series in terms of their rank. It should be mentioned that the class of finite rank time series contains the sum of the products of polynomials, exponentials and sinusoids. It should also be mentioned that any sine-wave time series (so-called sinusoid) with a frequency in the range (0,0.5) has rank 2, and the saw-tooth sinusoid with frequency 0.5 has rank 1. Subsequently, almost any periodic signal series with a finite number of addends has a finite rank. Any periodic time series definitely has a finite rank. Aperiodic time series such as white noise cannot have a finite rank. Note that the rank of the modulated periodic series may increase, but can stay finite. The only possible example of modulation that does not change the rank of the signal is the exponential modulation $\exp(\alpha t) = f^t$ with $f = e^\alpha$.

Trend signals have very varied and, as a rule, non-structured behaviour; in addition, the trend makes the main contribution towards the non-stationarity of the series. A typical trend that is a slowly varying component of the series can be precisely approximated by a series of finite rank. The list of slowly-varying series with simple SSA structure and small rank includes an exponential series (rank 1), a sinusoid with large period (rank 2), a linear series (rank 2) and polynomials of higher order (rank > 2) [6].

In the following section, we provide a general description and algorithm of the proposed approach to choosing r . In Section 6.3, the algorithm is applied to a number of simulated data with different structures. We use various signals such as linear trend, sequential exponential series, and sine series. This section also discuss the effect of noise level. Furthermore, despite the fact that the main aim is to select the optimal value of r we also briefly consider the selection of the value of L . The

applicability of the proposed approach is then examined using real ECG and EEG data. Section 6.4 provides the conclusions of this chapter.

6.2 Further development on the proposed approach to the identification of r

6.2.1 General description

In many cases, real-life time series have two main components, periodic (or quasi-periodic) and noise (aperiodic) components [6]. Thus, we assume that any time series can be written as $Y_N = \{y_t\}_{t=1}^N = S_N + E_N$, where S_N, E_N are the signal and noise components, respectively.

In order to separate the signal and noise components from each other, we first study the symmetry/skewness and kurtosis of the distribution of ζ_i for a white noise process; this type of noise will then be added to different signal series to see how the distribution of each eigenvalue would behave.

As mentioned earlier, the SSA technique consists of two main stages, decomposition and reconstruction; with each stage consisting of two compatible steps. The proposed approach is a novel step that can be used between the first and second stages of SSA to select the optimal value of r .

Consider a one-dimensional series Y_N , mapping this series into a Hankel matrix \mathbf{H} ($L \times K$), where $L \leq K$. Set $\mathbf{A} = \mathbf{H}\mathbf{H}^T$ and denoted by γ_i ($i = 1, \dots, L$) the eigenvalues of \mathbf{A} taken in the decreasing order of magnitude ($\gamma_1 \geq \dots \geq \gamma_L \geq 0$) and by U_1, \dots, U_L the orthonormal system of the eigenvectors of the matrix \mathbf{A} corresponding to these eigenvalues.

We mentioned in the previous chapter that the fundamental issue when studying the behaviour of the eigenvalues, γ_i , is that increasing the series length leads an increase

in the eigenvalues. This issue was solved by dividing \mathbf{A} by its trace, $\mathbf{B} = \mathbf{A}/tr(\mathbf{A})$. We also have pointed out that by utilising a simulation technique, the distribution of the eigenvalue of \mathbf{B} and its related forms can be used as a novel approach to the selection of the optimal value of r in SSA. In this regard, the main concept in proposing the present approach is separability, which characterises how well the signal and noise components can be separated from each other.

Let ζ_1, \dots, ζ_L denote the eigenvalues of matrix \mathbf{B} in decreasing order of magnitude ($1 \geq \zeta_1 \geq \dots \geq \zeta_L \geq 0$). In this step, we perform the simulation technique to obtain the distribution of ζ_i , so we can understand the behaviour of each eigenvalue, which can be useful for obtaining the optimal value of r .

In this chapter and the following ones, we focus on the third and fourth central measures moments of the distribution of ζ_i , which are the skewness (*Skew*) and kurtosis (*Kurt*) to identify the number of eigenvalues r corresponding to the signal component. Skewness is a measure of asymmetry of the data distribution, whilst kurtosis describes the distribution of observed data in terms of shape or peak. The coefficient of skewness of the distribution for each eigenvalue, $Skew(\zeta_i)$ and of kurtosis, $Kurt(\zeta_i)$ for an m simulation, can be calculated from the following formulas:

$$Skew(\zeta_i) = \frac{\frac{1}{m} \sum_{n=1}^m (\zeta_{i,n} - \bar{\zeta}_i)^3}{\left[\frac{1}{m-1} \sum_{n=1}^m (\zeta_{i,n} - \bar{\zeta}_i)^2 \right]^{3/2}}, \quad (6.1)$$

$$Kurt(\zeta_i) = \frac{\frac{1}{m} \sum_{n=1}^m (\zeta_{i,n} - \bar{\zeta}_i)^4}{\left[\frac{1}{m} \sum_{n=1}^m (\zeta_{i,n} - \bar{\zeta}_i)^2 \right]^2} - 3. \quad (6.2)$$

Furthermore, we consider the coefficient of variation, CV , which is defined as the ratio of the standard deviation $\sigma(\zeta_i)$ and $\bar{\zeta}_i$ which can be calculated mathematically:

$$CV(\zeta_i) = \frac{\sigma(\zeta_i)}{\bar{\zeta}_i}. \quad (6.3)$$

Additionally, the Spearman correlation ρ_s between the eigenvalues ζ_i and ζ_j ($i, j = 1, \dots, L$) is also calculated to enhance the results obtained by those measures:

$$\rho_s = \text{cor}(\zeta_i, \zeta_j) = 1 - \frac{6 \sum d_n^2}{m(m^2 - 1)}, \quad (6.4)$$

where $d_n = x_n - y_n$ ($n = 1, \dots, m$) is the difference between x_n and y_n which are the ranks of $\zeta_{i,n}$ and $\zeta_{j,n}$ respectively, and $\zeta_{i,n}$ is the n -th observation for the i -th eigenvalue (ζ_i), $\bar{\zeta}_i = \left(\sum_{n=1}^m \zeta_{i,n} \right) / m$.

The matrix of the absolute values of the Spearman correlation show a full decomposition of the trajectory matrix, and in this decomposition each eigenvalue is related to an elementary matrix component of the SVD. If the absolute value of ρ_s is close to zero, then the corresponding components are almost orthogonal, but if it is close to one, then the two components are far from being orthogonal and so it is difficult to separate them. Thus, if $\rho_s = 0$ between two reconstructed components, this means that these two reconstructed series are separable. The values of ρ_s between the eigenvalues for the white noise are approximately significant (refer to Chapter 3), which helps in the discrimination of the noise part.

The proposed criteria above split the eigenvalues into two groups, the first is corresponding to the signal and the second to the noise. Once r is obtained, then Eq. (1.3) can be written as

$$\mathbf{H} = \mathbf{S} + \mathbf{E}, \quad (6.5)$$

where $\mathbf{S} = \sum_{i=1}^r \mathbf{H}_i$ is the signal matrix and $\mathbf{E} = \sum_{i=r+1}^L \mathbf{H}_i$ is the noise matrix. At the final step, we use diagonal averaging to transform matrix \mathbf{S} into a new series of length N .

These measures of difference between the eigenvalues related to the signal and noise components can specify the cut-off point of separability; that is, it can determine the number of leading SVD components that are separated from the residual. Thus, the last cut-off point of separability between the signal and noise components obtained by the suggested measures, corresponds to the rank estimation.

In this chapter, we consider all the measures or criteria presented above. The following section gives a general example for the identification of the value of r based on the values of all the measures. In Section 6.2.2, we consider Y_N as a white noise series in order to have a prior idea and image of the pattern or behaviour of the values of $Skew(\zeta_i)$, $Kurt(\zeta_i)$, $CV(\zeta_i)$ and the correlation matrix between the eigenvalues of \mathbf{B} . This will enable us to identify the eigenvalues r related to the noise component, then to enhance the eigenvalues corresponding to the signal component for a signal model mixed with white noise. Furthermore, in Section 6.3 we will show the applicability of the proposed approach in determining r considering different real and simulated data with different values of signal to noise ratio. In addition, the optimality of the selection of L will also be discussed.

6.2.2 General example

6.2.2.1 White noise process

Let us first consider Y_N as a white noise process of length $N = 200$, and calculate the eigenvalues of \mathbf{B} , ζ_i ($i = 1, \dots, L$). Note that $\zeta_i = \gamma_i / \sum_{i=1}^L \gamma_i$ and ($1 \geq \zeta_1 \geq \zeta_2, \dots, \zeta_L \geq 0$). To study the asymptotical behaviour of the eigenvalues, the simulation for each case was repeated $m = 10^4$ times. Different values of L were used: $L = 10, 25, 50, 100$. Note that $\bar{\zeta}_i$ has a decreasing order for different values of L for the white noise process. The pattern of the mean and variance of ζ_i were studied in Chapter 4, and so we focus here on the values of $Skew$, $Kurt$, CV , and ρ_s .

Figure 6.1 illustrates the coefficient of skewness for ζ_i , for different values of L . It is obvious that the maximum value is obtained for the largest eigenvalue ζ_1 for different values of L except in the case $L = N/2$. Note that ζ_1 has a positive skewed distribution for different values of L . However, the distribution of the last eigenvalue can be right/left skewed or symmetric, whilst the middle eigenvalues have approximately symmetric distributions. The results or the pattern of the coefficient of kurtosis for ζ_i are similar to the pattern of the coefficient of skewness for ζ_i .

Remember that the pattern of the coefficient of variation of ζ_i has a U shape, as illustrated in Chapter 4.

Furthermore, the correlation between ζ_i and ζ_j is evaluated. Figure. 6.2 (left) shows the matrix of the correlation between ζ_i and ζ_j , for $L = 50$, considering the Spearman correlation coefficient $\rho_s(\zeta_i, \zeta_j)$, in a 20-grade grey scale from white to black corresponding to the values of correlations from -1 to 1. The results clearly show that the elements of the correlation matrix are highly correlated. In addition, for the correlation between eigenvalues, we can only consider the correlation between ζ_i and ζ_{i+1} ($i = 1, \dots, L-1$) for further purposes. The minimum value was observed between ζ_2 and ζ_3 ; see Figure. 6.2 (right). Since similar results emerged for other values of L , they are not reported here.

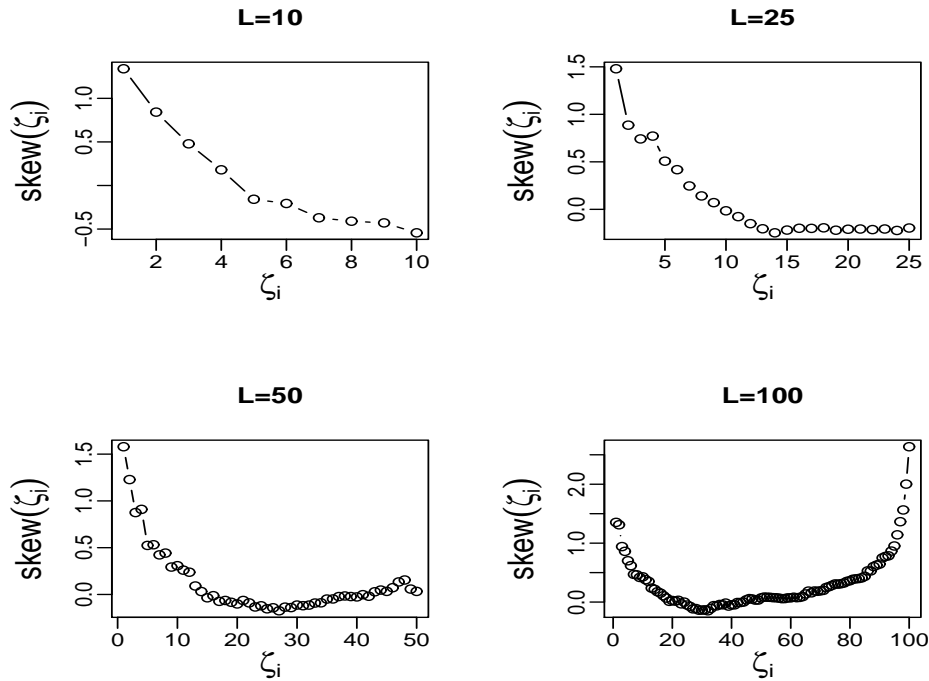


FIGURE 6.1: Skewness coefficient of ζ_i for different values of L for the white noise series.

6.2.2.2 Signal plus noise

In this part, we use five sine series with different amplitudes and frequencies and added white noise. One of the important features of the approach is that it does not

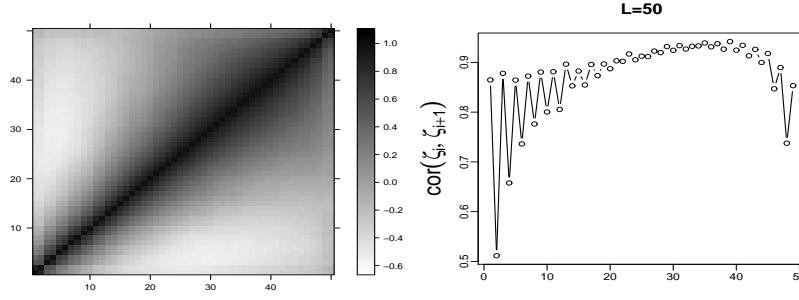


FIGURE 6.2: Matrix of correlation between ζ_i and $\zeta_j, i, j = (1, \dots, L)$ (left) and correlation coefficient between ζ_i and ζ_{i+1} (right) for the white noise series.

require the sinusoids to be separated from each other. Thus, the aim is to identify the value of r related to the whole sine series. The results obtained in the previous section help us to identify the patterns related to the noise component; we can then determine the required r for the signal series. In the following section, we will show that the proposed approach can be applied to various simulated and real series.

Let us now consider a time series $Y_N = \{y_t\}_{t=1}^N = S_N + E_N$ of length $N = 216$, where $S_N = \{s_t\}_{t=1}^N$,

$$s_t = 3.3 \left(s_t^{(1)} + s_t^{(2)} + s_t^{(3)} + s_t^{(4)} + s_t^{(5)} \right),$$

$$s_t^{(1)} = 0.8 \sin(2\pi t/12), s_t^{(2)} = 0.6 \sin(2\pi t/6), s_t^{(3)} = 0.4 \sin(2\pi t/4), s_t^{(4)} = 0.3 \sin(2\pi t/3), \\ s_t^{(5)} = 0.3 \sin(2\pi t/2.5)$$

is the signal and $E_N = \{\epsilon_t\}_{t=1}^N$ is a white noise process. The properties of the model usually depend on the signal-to-noise ratio $SNR = \sigma_s^2/\sigma_\epsilon^2$, here $SNR = 7$. Figure. 6.3 (left) represents this noisy time series.

As mentioned earlier, the distribution of ζ_i of matrix \mathbf{B} and its related forms will enable us to identify the value of r , and then separate and extract the signal component from the noisy time series. In SSA, the trend is any slowly varying component of the series, which does not contain cyclical/seasonal components. For trend extraction, a small window length should be enough [32], but this series does not include the trend. Thus, for this example we will not consider a very small window length.

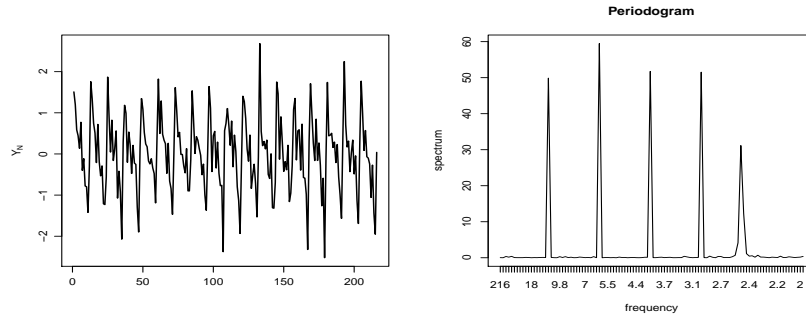
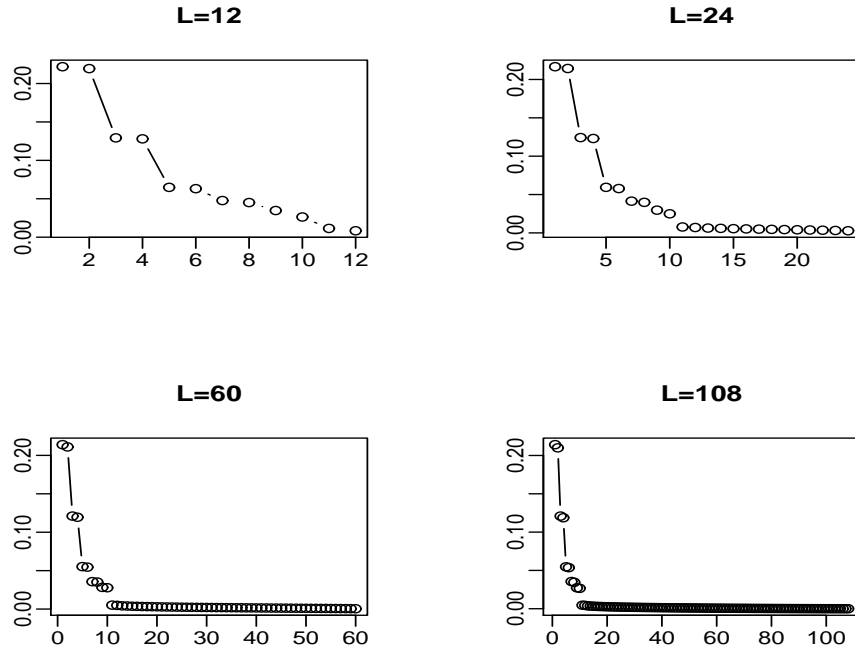


FIGURE 6.3: Five sine wave series with added noise (left) and its periodogram (right).

The signal here contains five sine waves with different frequencies. It is known that every harmonic component with a different frequency often produces two close eigenvalues (except for the frequency 0.5, which provides one eigenvalue with a saw-tooth singular vector) [112]. The periodogram analysis of the original series can help us to select the optimal L since it tells us which frequency must be considered. Figure. 6.3 (right) shows the periodograms of the original series. The information arising from the figure indicates that there are different frequencies, 12, 6, 4, 3, and 2.5. Thus, let us consider different values of window length, $L = 12, 24, 60, 108$.

Now, we can simulate m independent copies $E_{N,n}$ ($n = 1, \dots, m$) of the process E_N and find the distribution of ζ_i ($i = 1, \dots, L$) of matrix \mathbf{B} for m independent time series $Y_N = S_N + E_{N,n}$ for those values of window length. Figure. 6.4 depicts the plot of the logarithms of $\bar{\zeta}_i$ for the different values of L . It is obvious that there is a break in the eigenvalue spectra, particularly after ζ_{10} . Furthermore, it is known that a pure noise series produces a slowly decreasing sequence of eigenvalues. Moreover, it can be seen that there are five evident pairs with two almost equal eigenvalues, corresponding to an almost harmonic component of the series: component pairs 1-2, 3-4, 5-6, 7-8, 9-10 are related to harmonic components with specific periods 12, 6, 4, 3, and 2.5, respectively (for more information about periodograms of eigentriples and vectors, refer to [112]).

Remember that the result obtained in the previous example indicates that the largest eigenvalue for the white noise process has positive skewness distribution,

FIGURE 6.4: Logarithm of $\bar{\zeta}_i$ for different values of L for the five sine wave series.

which is closely associated with the second one. Generally, ζ_i has a strong relationship with ζ_{i+1} for the white noise process. Therefore, if $skew(\zeta_c)$ ($c \in \{1, \dots, L\}$) is the maximum, and the pattern for $skew(\zeta_c)$ to $skew(\zeta_L)$ has the same pattern, the same as emerged for the white noise, then the first $r = c - 1$ eigenvalues correspond to the signal and the rest to the noise. A similar procedure can be done using the coefficients of kurtosis and variation of ζ_i . Furthermore, if $\rho_s(\zeta_{c-1}, \zeta_c)$ is the minimum, and the pattern for the set $\{\rho_s(\zeta_i, \zeta_{i+1})\}_{i=c}^{L-1}$ is similar to what was observed for the white noise, then we select the first $r = c - 1$ eigenvalues for the signal and the rest for the noise component. Figure. 6.5 illustrates the result of $skew(\zeta_i)$ for different values of L . The result indicates that there are two clear shapes in the values of $skew(\zeta_i)$ for each window length. The pattern of the values for $skew(\zeta_{11})$ to $skew(\zeta_L)$ is similar to the pattern which emerged for the white noise process. It also can be seen that the maximum value of the skewness coefficient is observed for ζ_{11} for different values of L . Note that for large L , we are not concerned with the last eigenvalues, as they are obviously related to the noise component. Therefore, ζ_1 to ζ_{10} correspond to the signal component, and ζ_{11} to ζ_L belongs to the noise. The

results are also supported by the results obtained using the coefficients of kurtosis and variation of ζ_i (see Figure. 6.6)

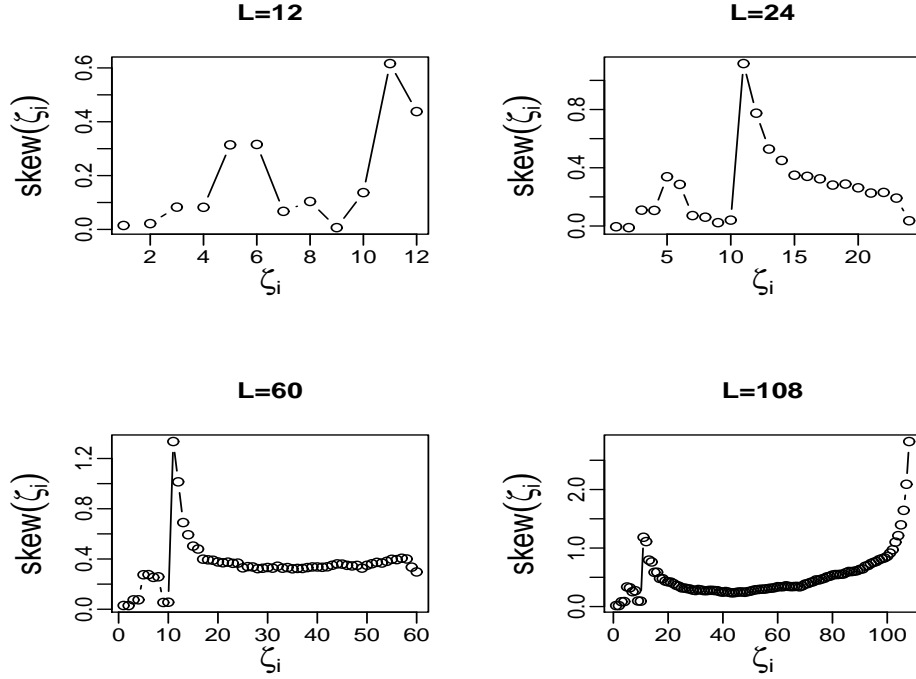


FIGURE 6.5: Coefficient of skewness of ζ_i for different values of L for the five sine wave series.

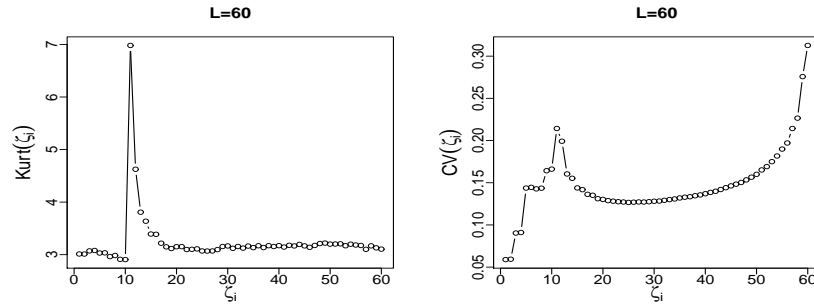


FIGURE 6.6: Coefficients of kurtosis and variation of ζ_i for $L = 60$ for the five sine series.

In addition to the previous results, the absolute value of the Spearman correlation between ζ_i and ζ_j is shown in Figure. 6.7 (left) for $L = 60$ (the same results are observed for other values of L). The correlation matrix enables us to distinguish and separate the different components from each other. Thus, the correlation matrix of ζ_i is considered as a new criterion for separability between these components.

If the absolute value of the correlation coefficient between ζ_i and ζ_j is small, then the corresponding components are almost orthogonal; however, if the value is large, then the corresponding series are far from being orthogonal and thus they are not neatly separable. It is clear that the signal can be separated from the noise since the top right pattern from the correlation matrix is similar to the pattern of the correlation matrix for the white noise obtained in the previous example.

In addition, Figure. 6.7 (right) shows the correlation between ζ_i and ζ_{i+1} . It is obvious that there are two different patterns in the figure, and the minimum value of ρ_s is achieved between ζ_{10} and ζ_{11} . Therefore, the results confirm that the cut-off point is from ζ_{11} . Note that we do not often consider the minimum value of ρ_s as a criterion for the cut-off point; the important consideration is to have two different shapes for $\rho_s(\zeta_i, \zeta_{i+1})$, and the second shape should be similar to the one observed for the white noise process.

To evaluate the results, the root mean square error (RMSE), and mean absolute error (MAE) between the original signal and reconstructed series were calculated. The results of these two measures confirm our results obtained that the value of r required for the signal series is 10 (see Figure. 6.8 for $L = 60$).

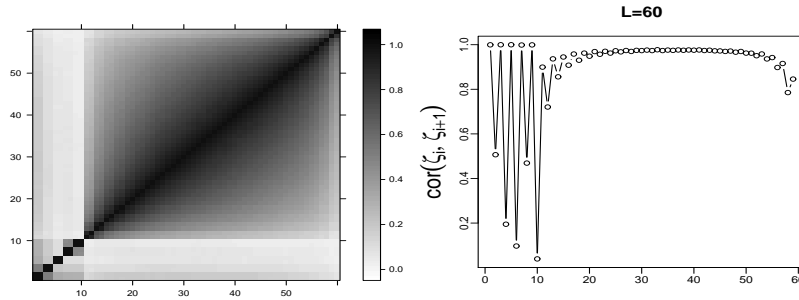


FIGURE 6.7: Correlation matrix of ζ_i for $L = 60$ (left) and the correlation coefficient between ζ_i and ζ_{i+1} (right) for the five sine series.

To examine this approach, we apply it to a real time series. However, the problem is how one can simulate this series. Here, we introduce a new approach for the simulation of a time series.

To simulate a value y_i , we generate a value from a uniform distribution with boundaries $y_i - a$ and $y_i + b$, where $a = |y_{i-1} - y_i|$ and $b = |y_i - y_{i+1}|$. This idea can be

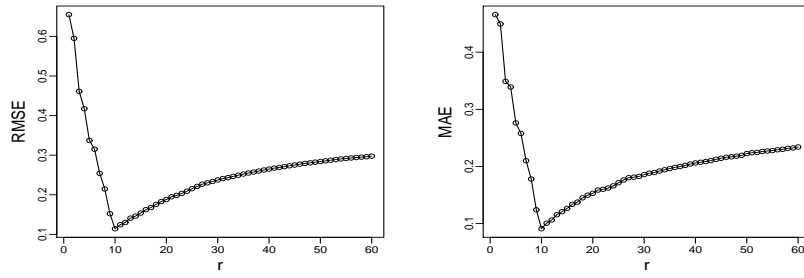


FIGURE 6.8: Root mean square error (RMSE) and mean absolute error (MAE) between the original signal and reconstructed series by r ($r = 1, \dots, 60$) for $L = 60$.

applied to any time series. It was applied to the simulated series Y_N in the previous example. The results obtained are similar. To assess the proposed simulation approach, four series were generated and simulated m times. The results are discussed in Section 6.3.

6.2.3 Algorithm

The algorithm is divided into two stages. At the first stage, the optimal value of r can be obtained for the separability between signal and noise, and at the second stage the free noise time series can be reconstructed.

6.2.3.1 Stage 1:

1. Transfer a one-dimensional time series $Y_N = (y_1, \dots, y_N)$ into the multi-dimensional series H_1, \dots, H_K with vectors $H_i = (y_i, \dots, y_{i+L-1})^T \in \mathbb{R}^L$, where $K = N - L + 1$, and the window length L is an integer such that $2 \leq L \leq N/2$. This step provides the trajectory matrix $\mathbf{H} = [H_1, \dots, H_K] = (x_{ij})_{i,j=1}^{L,K}$.
2. Compute matrix $\mathbf{B} = \mathbf{H}\mathbf{H}^T / \text{tr}(\mathbf{H}\mathbf{H}^T)$.
3. Compute the eigenvalues and eigenvectors of matrix \mathbf{B} and represent them in the form $\mathbf{B} = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^T$. Here, $\mathbf{\Sigma} = \text{diag}(\zeta_1, \dots, \zeta_L)$ is the diagonal matrix of

the eigenvalues of \mathbf{B} that has the order $(1 \geq \zeta_1 \geq \zeta_2, \dots, \zeta_L \geq 0)$ and $\mathbf{U} = (U_1, U_2, \dots, U_L)$ is the corresponding orthogonal matrix of the eigenvectors of \mathbf{B} .

4. Simulate the original series m times and calculate the eigenvalues for each series. We simulate y_i from a uniform distribution with boundaries $y_i - a$ and $y_i + b$, where $a = |y_{i-1} - y_i|$ and $b = |y_i - y_{i+1}|$.
5. Calculate the coefficient of skewness for each eigenvalue, $skew(\zeta_i)$. If $skew(\zeta_c)$ is the maximum, and the pattern for $skew(\zeta_c)$ to $skew(\zeta_L)$ has a similar pattern to what was emerged for the white noise, then select $r = c - 1$.
6. Calculate the coefficient of kurtosis for each eigenvalue, $kurt(\zeta_i)$. If $kurt(\zeta_c)$ is the maximum, then select $r = c - 1$.
7. Calculate the absolute values of the correlation matrix between the eigenvalues, and represent them in a 20-grade grey scale from white to black corresponding to the values of the correlations from 0 to 1. This can split the eigenvalues into two groups, from ζ_1 to ζ_r , which correspond to the signal, and the rest which correspond to the noise.
8. Calculate the absolute values of the correlation between ζ_i and ζ_{i+1} , and plot them in one figure. If $\rho_s(\zeta_{c-1}, \zeta_c)$ is the minimum, and the pattern for $\rho_s(\zeta_c, \zeta_{c+1})$ to $\rho_s(\zeta_{L-1}, \zeta_L)$ has the same pattern for the white noise, then choose $r = c - 1$.
9. Calculate the coefficient of variation, $CV(\zeta_i)$. This can split the eigenvalues into two groups, from ζ_1 to ζ_{c-1} , which correspond to the signal, and the rest, which almost have a U shape and which correspond to the noise component.

6.2.3.2 Stage 2

1. Use the number of the eigenvalues r obtained in the first stage to calculate the approximate signal matrix $\tilde{\mathbf{S}}$, that is $\tilde{\mathbf{S}} = \sum_{i=1}^r \mathbf{H}_i$, where $\mathbf{H}_i = \sqrt{\gamma_i} U_i V_i^T$, U_i and V_i stands for the left and right eigenvectors of the trajectory matrix.

2. Transition to the one dimensional series can now be achieved by averaging over the diagonals of the matrix $\tilde{\mathbf{S}}$.

6.3 Results

6.3.1 Simulated data

In the following examples, we have added a white noise process ϵ_t to four different signal series considering various signal to noise ratios:

- a. $y_t = \alpha_1 t + \epsilon_t$
- b. $y_t = \exp(\alpha_2 t) + \epsilon_t$
- c. $y_t = \alpha_3 \sin(2\pi t/12) + \epsilon_t$
- d. $y_t = \alpha_4 + \sin(2\pi t/6) + \sin(2\pi t/12) + \epsilon_t$,

where $t = (1, \dots, N)$, $\alpha_1 = 0.062$, $\alpha_2 = 0.011$, $\alpha_3 = 6.31$, $\alpha_4 = 0.078$ and ϵ_t is a Gaussian white noise process with variance 1. The signal to noise ratios for the above cases are $SNR = 15, 7, 20, 24$, respectively. For each series, a $m = 10000$ time series was simulated using the proposed simulation approach in Section 6.2.2, for a fixed $L = 12$ and $N = 216$. It is obvious that the number of eigenvalues needed to reconstruct the signal for these cases are 1, 1, 2, 5, respectively. Table 6.1 represents the skewness coefficient of ζ_i ($i = 1, \dots, L$) for all cases. As mentioned earlier, the maximum value of the skewness coefficient is the indicator of the start of the noise. It is clear that the maximum skewness coefficients of ζ_i for the four cases are for $\zeta_{c=2}$, $\zeta_{c=2}$, $\zeta_{c=3}$, $\zeta_{c=6}$, respectively. Thus, the number of eigenvalues required to extract the signal for those cases are 1, 1, 2, 5 as $r = c - 1$. Similar results emerged by using the values of $Kurt$ and CV (see Tables 6.2 and 6.3). Moreover, the correlation between ζ_i and ζ_{i+1} for the four cases is considered. Table. 6.4 represents the correlation between ζ_i and ζ_{i+1} . For the correlation coefficient, we need the minimum value of ρ_s between ζ_{c-1} and ζ_c as an indicator for the cut-off

point, with a condition that the pattern of the set $\{\rho_s(\zeta_i, \zeta_{i+1})\}_{i=c}^{L-1}$ is similar to that obtained for the white noise process. The results are similar to what emerged with the coefficient of skewness, and confirm that the proposed approach works properly. The results of RMSE and MAE also confirm that the proposed approach works efficiently (see Tables 6.5 and 6.6). Thus, the proposed simulation approach will also be used for simulating real time series in section 6.3.4.

TABLE 6.1: Skewness coefficient of ζ_i for cases a, ..., d.

Case	ζ_1	ζ_2	ζ_3	ζ_4	ζ_5	ζ_6	ζ_7	ζ_8	ζ_9	ζ_{10}	ζ_{11}	ζ_{12}
a	-0.06	0.362*	0.26	0.19	0.10	0.16	-0.09	-0.06	-0.03	-0.04	0.06	0.002
b	-0.09	0.28*	0.25	0.22	0.25	0.12	0.18	0.10	0.10	0.05	-0.05	0.06
c	0.24	0.19	1.10*	0.72	0.37	0.25	0.09	-0.02	-0.14	-0.16	-0.15	-0.42
d	0.002	0.18	0.15	-0.06	-0.04	0.47*	0.29	0.22	0.19	0.13	0.05	-0.06

Note :*, represents the maximum value of $skew(\zeta_i)$.

TABLE 6.2: Kurtosis coefficient of ζ_i for cases a, ..., d.

Case	ζ_1	ζ_2	ζ_3	ζ_4	ζ_5	ζ_6	ζ_7	ζ_8	ζ_9	ζ_{10}	ζ_{11}	ζ_{12}
a	3.08	4.16*	3.66	3.45	3.12	3.07	3.02	3.08	3.08	3.11	3.03	3.10
b	3.08	4.12*	3.53	3.27	3.29	3.10	3.07	3.10	3.01	3.05	3.06	3.12
c	3.07	3.04	4.03*	3.32	3.20	3.28	3.13	3.11	3.11	3.13	3.09	3.10
d	3.04	3.24	3.26	3.05	3.03	4.00*	3.22	3.15	3.02	2.95	2.98	3.02

Note :*, represents the maximum value of $kurt(\zeta_i)$.

TABLE 6.3: Variation coefficient of ζ_i for cases a, ..., d.

Case	ζ_1	ζ_2	ζ_3	ζ_4	ζ_5	ζ_6	ζ_7	ζ_8	ζ_9	ζ_{10}	ζ_{11}	ζ_{12}
a	0.002	0.13*	0.12	0.113	0.112	0.110	0.112	0.114	0.119	0.12	0.14	0.16
b	0.004	0.14*	0.13	0.116	0.115	0.113	0.114	0.116	0.12	0.125	0.14	0.16
c	0.00850	0.00855	0.146*	0.132	0.121	0.1180	0.1170	0.1174	0.122	0.128	0.14	0.16
d	0.002	0.119	0.122	0.130	0.133	0.152*	0.145	0.133	0.1376	0.1379	0.155	0.170

Note :*, represents the peak value of $CV(\zeta_i)$.

TABLE 6.4: Correlation coefficient between ζ_i and ζ_{i+1} for cases a, ..., d.

Case	$\rho(\zeta_i, \zeta_{i+1})$											
a	0.52*	0.85	0.73	0.60	0.91	0.78	0.93	0.78	0.93	0.72	0.98	
b	0.52*	0.78	0.70	0.71	0.81	0.85	0.74	0.83	0.80	0.76	0.90	
c	0.72	0.41*	0.71	0.53	0.74	0.70	0.75	0.74	0.76	0.65	0.69	
d	0.72	0.99	0.04	0.99	0.03*	0.88	0.70	0.85	0.80	0.76	0.78	

Note :*, represents the minimum value of $\rho(\zeta_i, \zeta_{i+1})$.

TABLE 6.5: Root mean square error between the original and reconstructed series by $(r = 1, \dots, 12)$ for cases a, ..., d.

Case	$r = 1$	$r = 2$	$r = 3$	$r = 4$	$r = 5$	$r = 6$	$r = 7$	$r = 8$	$r = 9$	$r = 10$	$r = 11$	$r = 12$
a	0.07*	0.11	0.15	0.18	0.20	0.22	0.24	0.26	0.28	0.29	0.30	0.32
b	0.05*	0.09	0.13	0.15	0.17	0.20	0.22	0.24	0.25	0.27	0.28	0.29
c	1.20	0.31*	0.41	0.52	0.60	0.65	0.71	0.75	0.83	0.91	0.95	0.99
d	1	0.78	0.71	0.37	0.16*	0.17	0.20	0.22	0.24	0.26	0.27	0.29

Note :*, represents the minimum value of RMSE.

6.3.2 The effect of noise level

In addition to the previous results, we also consider different signal to noise ratios for each case to reach a better understanding of the effect of noise reduction using the

TABLE 6.6: Mean absolute error between the original and reconstructed series by $(r = 1, \dots, 12)$ for cases a, ..., d.

Case	$r = 1$	$r = 2$	$r = 3$	$r = 4$	$r = 5$	$r = 6$	$r = 7$	$r = 8$	$r = 9$	$r = 10$	$r = 11$	$r = 12$
a	0.06*	0.09	0.11	0.14	0.16	0.17	0.19	0.20	0.22	0.23	0.24	0.25
b	0.04*	0.07	0.10	0.12	0.14	0.16	0.18	0.19	0.20	0.21	0.22	0.24
c	1.60	0.28*	0.34	0.42	0.50	0.53	0.56	0.61	0.64	0.72	0.75	0.80
d	0.76	0.63	0.63	0.32	0.13*	0.14	0.16	0.18	0.19	0.20	0.22	0.23

Note :*, represents the minimum value of MAE.

proposed approach. As we mentioned the properties of the model usually depend on the signal-to-noise ratio (SNR). Here the SNR is the ratio of variance of the noise free series (signal) to variance of noise $SNR = \sigma_s^2 / \sigma_\epsilon^2$ (for more information about the definition of SNR, see [250]). The results are provided for cases a and c, as similar results are observed for the other two cases. Figures. 6.9 and 6.10 show $Skew(\zeta_i)$ for cases a and c considering different values of SNR. For example, Figure. 6.9 shows $Skew$ for the case a, where we have only a trend component. As shown in the figure, the maximum value of $Skew$ is observed for $\zeta_{c=2}$ for different values of SNR, which confirms that $r = 1$. Figure. 6.10 shows $Skew$ for the case c, where we have a harmonic component. It can be seen from the figure that $r = 2$ based on the results of $Skew$ for different values of SNR. This result confirms that the new approach works for any series that is mixed with a low or high noise level.

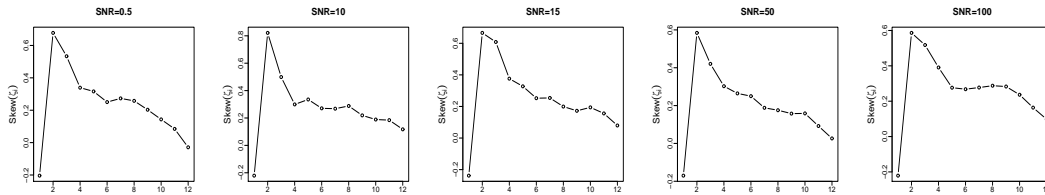


FIGURE 6.9: Value of skewness coefficient of eigenvalues for case a, considering different values of SNR.

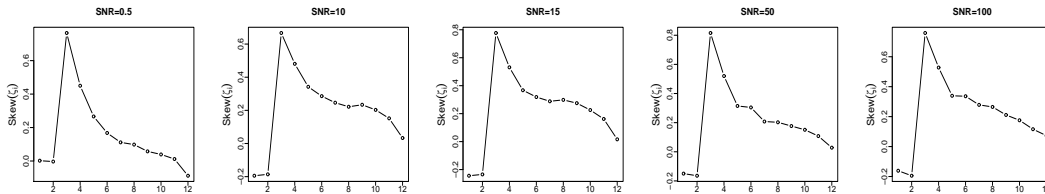


FIGURE 6.10: Value of skewness coefficient of eigenvalues for case c, considering different values of SNR.

6.3.3 Selection of L

As mentioned earlier, we will briefly consider the selection of window length. Here we consider different values of L , namely 12, 24, 60, 108 to analyse each series considered in section 6.3. The value of SNR here is fixed (SNR=14) for all series, and the number of simulations is also 10,000. In this case, based on the values of window length on SVD of the trajectory matrix, we have 12, 24, 60, 108 eigenvalues or eigentriples for each case, ordered by their contributions into the decomposition stage. However, as we already know the maximum number of eigenvalues needed was five for case d, as it consists of a trend and two sine series, we will therefore only provide the results for the first 12 eigenvalues for comparison between these four cases. First we need to identify the set of leading eigenvalues for each value of L in each case. All measures are considered in this section to help in identifying the optimal value of L . The results are compared with the results of the w -correlations and evaluated using RMSE and MAE.

Tables 6.7, 6.8, 6.9, and 6.10 represent the results of the coefficient of skewness, kurtosis, variation for each eigenvalue and the results of the correlation between ζ_i and ζ_{i+1} . As shown by the results, for the sequential exponential series b considering different values of L , the maximum values of $Skew$, $Kurt$, CV are observed for $\zeta_{c=2}$, and the minimum value of ρ_s is obtained between $\zeta_{c-1=1}$ and $\zeta_{c=2}$ for each value of L ; this indicates that there is no effect of L and thus $r = c - 1 = 1$. A similar result emerged for sine series c. The value of r is 2 for different values of L because the maximum values of $Skew$, $Kurt$, and CV are observed for $\zeta_{c=3}$, and the minimum value of ρ_s is obtained between $\zeta_{c-1=2}$ and $\zeta_{c=3}$. However, for the linear trend series a, the value of r is 1 for $L = 12$ and $L = 24$, whilst it is 2 for $L = 60$ and $L = 108$ based on all the measures. This means that the rank or r increases from 1 to 2 with the increase in L , but the increase is only by one from small to large window length ($L = N/2$), and the rank is still finite. A similar result are obtained for series d because it includes a trend component with two sine series. For $L = 12$ and $L = 24$, the the maximum values of $Skew$, $Kurt$, and CV is obtained for $\zeta_{c=6}$, and the minimum value of ρ_s is observed between $\zeta_{c-1=5}$ and $\zeta_{c=6}$; this indicates that

$r = 5$. However, for $L = 60$ and $L = 108$, the value of r increases from 5 to 6 based on all the criteria.

Let us now compare the results with the results of the w -correlations. Table 6.11 represents the results obtained from the w -correlations for the four simulated series. It can be seen from the table that the results of the w -correlations for cases a, b, and c coincide with the results obtained by the above criteria for different values of window length. It can also be seen that for case d, the results of the w -correlations only coincide with those criteria for $L = N/2 = 108$. It is obvious that the results based on *Skew*, *Kurt*, *CV*, and ρ_s are more accurate than those obtained by the w -correlations for small window length, particularly for case d, where a linear trend is included in the series.

We now aim to fix r and select the optimal value of L based on the maximum values of the *Skew*, *Kurt*, *CV* of ζ_i and the minimum value of ρ_s between ζ_i and ζ_{i+1} , w -correlations, RMSE, and MAE. The values of r are fixed for the four cases based on the above results, being $r = 1, 2$ for series a, $r = 1$ for series b, $r = 2$ for series c, and $r = 5, 6$ for series d. Tables 6.12, 6.13, 6.14, and 6.15 show the results. As appears from these results, the best reconstructed signals for these series, considering all the different criteria, are obtained using $L = N/2 = 108$, and $r = 2, 1, 2, 6$ for the four series respectively. The results indicate that the optimal value of window length is half of the series for all the considered series. Figure 6.11 depicts the the result of the reconstructed series, which is obtained by using $L = N/2$ and eigentriples $1 - r$ ($r = 2, 1, 2, 6$) for a, b, c, and d series, respectively. The blue and the black lines correspond to the reconstructed series and the original series. It is obvious that the reconstructed series has been obtained precisely.

6.3.4 Real data

We consider two real time series, namely an electrocardiogram (ECG) and an electroencephalogram (EEG) series. The window lengths used to decompose these series were 460, and 200, respectively. The aim here is to find the value of r for each series.

TABLE 6.7: Skewness coefficient of ζ_i for cases a, ..., d and different values of L .

	ζ_1	ζ_2	ζ_3	ζ_4	ζ_5	ζ_6	ζ_7	ζ_8	ζ_9	ζ_{10}	ζ_{11}	ζ_{12}
L	Case a											
12	0.270	1.139*	0.742	0.424	0.279	0.027	-0.111	-0.159	-0.197	-0.249	-0.320	-0.391
24	0.178	0.905*	0.794	0.540	0.520	0.411	0.298	0.256	0.151	0.093	0.045	-0.002
60	0.063	0.271	1.168*	1.078	0.850	0.841	0.612	0.600	0.461	0.462	0.360	0.337
108	-0.007	0.326	1.219*	1.151	0.899	0.912	0.699	0.674	0.569	0.559	0.430	0.421
L	Case b											
12	0.191	1.078*	0.731	0.448	0.262	0.062	-0.056	-0.123	-0.184	-0.240	-0.269	-0.331
24	0.186	1.185*	0.955	0.595	0.544	0.423	0.369	0.274	0.236	0.118	0.066	0.008
60	0.068	1.416*	1.156	0.769	0.756	0.565	0.561	0.509	0.502	0.418	0.419	0.335
108	-0.024	1.532*	1.252	0.877	0.845	0.611	0.637	0.553	0.564	0.486	0.493	0.391
L	Case c											
12	0.233	0.170	1.007*	0.629	0.304	0.210	0.072	-0.015	-0.128	-0.149	-0.139	-0.322
24	0.243	0.222	1.172*	0.893	0.545	0.475	0.321	0.225	0.172	0.108	0.060	0.043
60	0.124	0.122	1.363*	1.124	0.821	0.744	0.595	0.567	0.415	0.397	0.281	0.267
108	0.026	0.024	1.448*	1.205	0.850	0.821	0.704	0.664	0.502	0.491	0.381	0.368
L	Case d											
12	-0.012	0.312	0.314	0.032	0.039	0.835*	0.487	0.218	0.090	-0.024	-0.087	-0.242
24	-0.015	0.360	0.357	0.065	0.069	0.909*	0.804	0.431	0.424	0.256	0.067	0.148
60	-0.046	0.481	0.385	0.244	0.110	0.123	1.211*	1.134	0.815	0.779	0.623	0.602
108	-0.065	0.466	0.395	0.201	0.118	0.078	1.262*	1.173	0.936	0.842	0.674	0.678

Note :*, represents the maximum value of $skew(\zeta_i)$.

TABLE 6.8: Kurtosis coefficient of ζ_i for cases a, ..., d and different values of L .

	ζ_1	ζ_2	ζ_3	ζ_4	ζ_5	ζ_6	ζ_7	ζ_8	ζ_9	ζ_{10}	ζ_{11}	ζ_{12}
L	Case a											
12	3.229	5.426*	3.854	3.380	3.359	3.244	3.174	3.055	2.999	3.051	3.026	3.141
24	3.091	4.330*	4.022	3.467	3.487	3.347	3.114	3.138	3.139	3.154	3.071	3.001
60	3.010	3.092	5.323*	4.822	4.259	4.199	3.668	3.623	3.353	3.401	3.286	3.241
108	3.041	3.046	5.571*	5.002	4.445	4.519	3.903	3.883	3.674	3.683	3.324	3.338
L	Case b											
12	3.109	5.270*	3.983	3.569	3.292	3.055	3.081	3.105	2.999	3.026	3.040	3.112
24	3.154	5.741*	4.661	3.616	3.566	3.394	3.285	3.304	3.307	3.102	3.021	3.065
60	3.031	7.516*	5.363	4.130	4.061	3.494	3.518	3.497	3.509	3.375	3.401	3.230
108	2.956	8.535*	5.754	4.598	4.375	3.671	3.774	3.530	3.517	3.440	3.507	3.271
L	Case c											
12	2.996	3.012	5.039*	3.864	3.295	3.215	3.141	3.146	3.024	2.991	2.975	3.135
24	2.998	3.012	5.685*	4.458	3.573	3.458	3.237	3.141	3.182	3.109	3.076	3.028
60	2.992	2.950	6.676*	5.159	4.341	4.021	3.373	3.635	3.393	3.348	3.167	3.144
80	2.981	2.969	7.266*	5.557	4.373	4.268	3.995	3.971	3.587	3.539	3.257	3.237
L	Case d											
12	2.959	3.278	3.280	2.919	2.916	4.255*	3.431	3.109	3.027	3.042	2.974	3.079
24	2.966	3.349	3.348	2.940	2.945	4.548*	4.230	3.367	3.468	3.084	3.274	3.133
60	2.935	3.470	3.397	2.938	2.948	3.020	5.593*	5.205	4.286	4.075	3.738	3.724
108	2.936	3.557	3.380	3.056	2.945	3.048	5.938*	5.280	4.848	4.255	3.782	3.819

Note :*, represents the maximum value of $kurt(\zeta_i)$.

TABLE 6.9: Variation coefficient of ζ_i for cases a, ..., d and different values of L .

L	ζ_1	ζ_2	ζ_3	ζ_4	ζ_5	ζ_6	ζ_7	ζ_8	ζ_9	ζ_{10}	ζ_{11}	ζ_{12}
	Case a											
12	0.0005	0.1005*	0.087	0.065	0.063	0.059	0.062	0.065	0.074	0.082	0.107	0.125
24	0.0006	0.136*	0.117	0.098	0.084	0.078	0.071	0.069	0.066	0.066	0.065	0.066
60	0.0010	0.163	0.175*	0.169	0.123	0.121	0.100	0.099	0.089	0.089	0.082	0.081
108	0.0015	0.200	0.202*	0.199	0.143	0.142	0.119	0.118	0.106	0.06	0.098	0.097
	Case b											
12	0.0009	0.102*	0.089	0.068	0.066	0.061	0.064	0.066	0.073	0.081	0.104	0.122
24	0.001	0.130*	0.119	0.086	0.082	0.072	0.070	0.066	0.066	0.063	0.064	0.064
60	0.002	0.178*	0.170	0.120	0.119	0.100	0.101	0.090	0.089	0.084	0.083	0.079
108	0.002	0.207*	0.201	0.142	0.142	0.119	0.119	0.108	0.107	0.100	0.099	0.094
	Case c											
12	0.00328	0.0032	0.116*	0.097	0.079	0.076	0.075	0.077	0.082	0.090	0.112	0.136
24	0.00326	0.0033	0.141*	0.123	0.093	0.088	0.079	0.077	0.074	0.073	0.072	0.073
60	0.0040	0.0041	0.187*	0.173	0.127	0.122	0.106	0.103	0.094	0.092	0.087	0.086
108	0.0046	0.0047	0.214*	0.202	0.147	0.143	0.124	0.121	0.111	0.109	0.102	0.101
	Case d											
12	0.0026	0.118	0.120	0.126	0.129	0.130*	0.111	0.094	0.100	0.101	0.124	0.140
24	0.0028	0.134	0.137	0.145	0.147	0.149*	0.121	0.096	0.090	0.084	0.080	0.078
60	0.0032	0.143	0.155	0.153	0.170	0.147	0.180*	0.174	0.126	0.123	0.105	0.103
108	0.0036	0.131	0.163	0.136	0.182	0.160	0.208*	0.205	0.147	0.145	0.123	0.122

Note :*, represents the peak value of $CV(\zeta_i)$.

1. ECG:

An electroencephalogram (ECG) records the electrical activity of the heart.

The heart produces tiny electrical impulses which spread through the heart

TABLE 6.10: Correlation coefficient between ζ_i and ζ_{i+1} for cases a, . . . , d and considering different values of L .

$\rho(\zeta_i, \zeta_{i+1})$											
L	Case a										
12	0.278*	0.705	0.502	0.661	0.634	0.663	0.730	0.672	0.797	0.602	0.812
24	0.094*	0.276	0.820	0.598	0.850	0.695	0.859	0.755	0.856	0.784	0.868
60	0.405	0.029*	0.927	0.591	0.932	0.736	0.929	0.795	0.931	0.842	0.932
108	0.602	0.021*	0.945	0.595	0.951	0.742	0.953	0.808	0.950	0.846	0.952
Case b											
12	0.354*	0.765	0.451	0.737	0.597	0.729	0.696	0.708	0.763	0.625	0.792
24	0.344*	0.861	0.546	0.862	0.680	0.855	0.752	0.854	0.792	0.858	0.828
60	0.315*	0.920	0.599	0.930	0.738	0.931	0.802	0.931	0.853	0.934	0.866
108	0.303*	0.940	0.608	0.944	0.745	0.948	0.815	0.949	0.854	0.952	0.879
Case c											
12	0.790	0.466*	0.691	0.565	0.727	0.706	0.746	0.744	0.752	0.674	0.710
24	0.921	0.380*	0.811	0.588	0.837	0.742	0.858	0.796	0.868	0.838	0.872
60	0.973	0.359*	0.894	0.627	0.908	0.763	0.917	0.821	0.927	0.860	0.925
108	0.984	0.353*	0.918	0.629	0.926	0.758	0.934	0.824	0.939	0.869	0.939
Case d											
12	0.812	0.982	0.520	0.989	0.078*	0.767	0.567	0.727	0.719	0.639	0.772
24	0.824	0.993	0.496	0.995	0.004*	0.492	0.758	0.669	0.824	0.748	0.838
60	0.813	0.990	0.510	0.981	0.142	0.030*	0.924	0.598	0.928	0.743	0.925
108	0.727	0.784	0.674	0.752	0.658	0.001*	0.945	0.613	0.946	0.750	0.947

Note :*, represents the minimum value of $\rho(\zeta_i, \zeta_{i+1})$.

TABLE 6.11: Value of w -correlations of the signal reconstruction for different values of L and r for cases a, . . . , d.

L	$r = 1$	$r = 2$	$r = 3$	$r = 4$	$r = 5$	$r = 6$	$r = 7$	$r = 8$	$r = 9$	$r = 10$	$r = 11$
Case a											
12	0.0054*	0.0162	0.0160	0.0173	0.0203	0.0306	0.0267	0.0268	0.0281	0.0213	0.0305
24	0.0055*	0.0113	0.0125	0.0102	0.0147	0.0162	0.0197	0.0212	0.0201	0.0224	0.0239
60	0.0127	0.0011*	0.0051	0.0038	0.0065	0.0069	0.0087	0.0097	0.0108	0.0112	0.0145
108	0.0070	0.0007*	0.0038	0.0021	0.0046	0.0036	0.0069	0.0096	0.0106	0.0102	0.0092
Case b											
12	0.0053*	0.0186	0.0311	0.0357	0.0277	0.0449	0.0463	0.0526	0.0683	0.0661	0.0516
24	0.0021*	0.0116	0.0163	0.0176	0.0138	0.0197	0.0291	0.0328	0.0365	0.0344	0.0336
60	0.0005*	0.0061	0.0032	0.0082	0.0074	0.0114	0.0153	0.0165	0.0194	0.0210	0.0214
108	0.0004*	0.0054	0.0016	0.0061	0.0063	0.0101	0.0106	0.0133	0.0145	0.0142	0.0121
Case c											
12	0.8972	0.0140*	0.0415	0.0525	0.0565	0.0361	0.0635	0.0641	0.0614	0.0866	0.1147
24	0.8850	0.0057*	0.0202	0.0277	0.0296	0.0233	0.0303	0.0500	0.0507	0.0562	0.0546
60	0.8798	0.0017*	0.0098	0.0056	0.0128	0.0172	0.0228	0.0243	0.0286	0.0309	0.0314
108	0.8793	0.0006*	0.0074	0.0022	0.0084	0.0088	0.0141	0.0149	0.0185	0.0197	0.0199
Case d											
12	0.0264	0.0211	0.0108*	0.0206	0.0122	0.0256	0.0288	0.0268	0.0293	0.0221	0.0313
24	0.0123	0.0314	0.0197	0.0066*	0.0138	0.0136	0.0129	0.0148	0.0195	0.0237	0.0257
60	0.0108	0.0365	0.0354	0.0304	0.0160	0.0024*	0.0067	0.0053	0.0080	0.0070	0.0104
108	0.0063	0.0267	0.0368	0.0212	0.0207	0.0012*	0.0045	0.0025	0.0056	0.0066	0.0095

Note :*, represents the minimum value of w -correlations.

muscle to make the heart contract. However, the heart sound may contain various abnormal components. The contribution of this work is to select the correct subspace of the signal component and remove the undesired component. The ECG data used in this paper was described in the previous chapter and downloaded from [248], and is shown in Figure. 6.12 (left). Remember that the signal is abnormal, and the distribution of this signal is not normal (see Figure. 6.13 (left)).

The matrix of the absolute values of the Spearman correlation of ζ_i is shown in Figure. 6.12 (middle). It is very clear that the eigenvalues are split into two groups, from the first to the 150-th and the rest, which decomposes the

TABLE 6.12: Simulated series: value of *Skew*, *Kurt*, *CV*, ρ_s , w-correlation, MAE, and RMSE of the signal reconstruction step for different values of L and r for case a.

Case a				
$r = 1$				
	$L = 12$	$L = 24$	$L = 60$	$L = 108$
RMSE	0.2810	0.2113	0.4725	0.7183
MAE	0.2319	0.1694	0.3035	0.5356
w-correlation	0.0054	0.0055	0.0127	0.0070
$Skew(\zeta_2)$	1.1392	0.9051	0.2715	0.3258
$Kurt(\zeta_2)$	5.4263	4.3210	3.0923	3.0459
$CV(\zeta_2)$	0.1005	0.1360	0.1629	0.1998
$\rho(\zeta_1, \zeta_2)$	0.2779	0.0937	0.4054	0.6020
$r = 2$				
	$L = 12$	$L = 24$	$L = 60$	$L = 108$
RMSE	0.3887	0.2992	0.2235	0.2110*
MAE	0.3001	0.2451	0.1760	0.1659*
w-correlations	0.0162	0.0113	0.0012	0.0007*
$Skew(\zeta_3)$	0.7417	0.7938	1.1686	1.2187*
$Kurt(\zeta_3)$	3.8544	4.0224	5.3231	5.5706*
$CV(\zeta_3)$	0.0869	0.1169	0.1751	0.2023*
$\rho(\zeta_2, \zeta_3)$	0.7054	0.2765	0.0286	0.0209*

Note :*, represents the optimal value of L and r for case a based on all the considered criteria.

TABLE 6.13: Simulated series: value of *Skew*, *Kurt*, *CV*, ρ_s , w-correlation, MAE, and RMSE of the signal reconstruction step for different values of L and r for case b.

Case b				
$r = 1$				
	$L = 12$	$L = 24$	$L = 60$	$L = 108$
RMSE	0.2141	0.1473	0.1106	0.0914*
MAE	0.1771	0.1312	0.0864	0.0677*
w-correlation	0.0053	0.0021	0.0005	0.0004*
$Skew(\zeta_2)$	1.0775	1.1853	1.4163	1.5317*
$Kurt(\zeta_2)$	5.2701	5.7414	7.5161	8.5354*
$CV(\zeta_2)$	0.1019	0.1297	0.1778	0.2073*
$\rho(\zeta_1, \zeta_2)$	0.3538	0.3444	0.3152	0.3028*

Note :*, represents the optimal value of L and r for case b based on all the considered criteria.

trajectory matrix into almost orthogonal blocks, with the first block corresponding to the approximated version of the ECG series and the second block

TABLE 6.14: Simulated series: value of *Skew*, *Kurt*, *CV*, ρ_s , *w*-correlation, MAE, and RMSE of the signal reconstruction step for different values of L and r for case c.

Case c				
$r = 2$				
	$L = 12$	$L = 24$	$L = 60$	$L = 108$
RMSE	0.3830	0.2666	0.1694	0.1425*
MAE	0.3232	0.2193	0.1369	0.1202*
<i>w</i> -correlation	0.0140	0.0057	0.0016	0.0006*
<i>Skew</i> (ζ_3)	1.0070	1.1723	1.3631	1.4481*
<i>Kurt</i> (ζ_3)	5.0388	5.6854	6.6760	7.2655*
<i>CV</i> (ζ_3)	0.1156	0.1414	0.1868	0.2141*
$\rho(\zeta_2, \zeta_3)$	0.4659	0.3800	0.3585	0.3529*

Note :*, represents the optimal value of L and r for case c based on all the considered criteria.

TABLE 6.15: Simulated series: value of *Skew*, *Kurt*, *CV*, ρ_s , *w*-correlation, MAE, and RMSE of the signal reconstruction step for different values of L and r for case d.

Case d				
$r = 5$				
	$L = 12$	$L = 24$	$L = 60$	$L = 108$
RMSE	0.6219	0.3550	0.4339	0.5123
MAE	0.4961	0.2852	0.3146	0.3863
<i>w</i> -correlation	0.0122	0.0138	0.0160	0.0207
<i>Skew</i> (ζ_6)	0.8346	0.9086	0.1233	0.0784
<i>Kurt</i> (ζ_6)	4.2554	4.5476	3.0201	3.0482
<i>CV</i> (ζ_6)	0.1222	0.1302	0.1473	0.1595
$\rho(\zeta_5, \zeta_6)$	0.0785	0.0041	0.1420	0.6579
$r = 6$				
	$L = 12$	$L = 24$	$L = 60$	$L = 108$
RMSE	0.6792	0.4179	0.2070	0.1939*
MAE	0.5434	0.3319	0.1636	0.1588*
<i>w</i> -correlations	0.0256	0.0136	0.0024	0.0012*
<i>Skew</i> (ζ_7)	0.4867	0.8037	1.2112	1.2617*
<i>Kurt</i> (ζ_7)	3.4311	4.2303	5.5927	5.9378*
<i>CV</i> (ζ_7)	0.1107	0.1206	0.1798	0.0208*
$\rho(\zeta_6, \zeta_7)$	0.7671	0.4925	0.0305	0.0011*

Note :*, represents the optimal value of L and r for case d based on all the considered criteria.

corresponding to the noise part. Furthermore, Figure. 6.12 (right) indicates that there are two different sequences for the result of $skew(\zeta_i)$, from $skew(\zeta_1)$

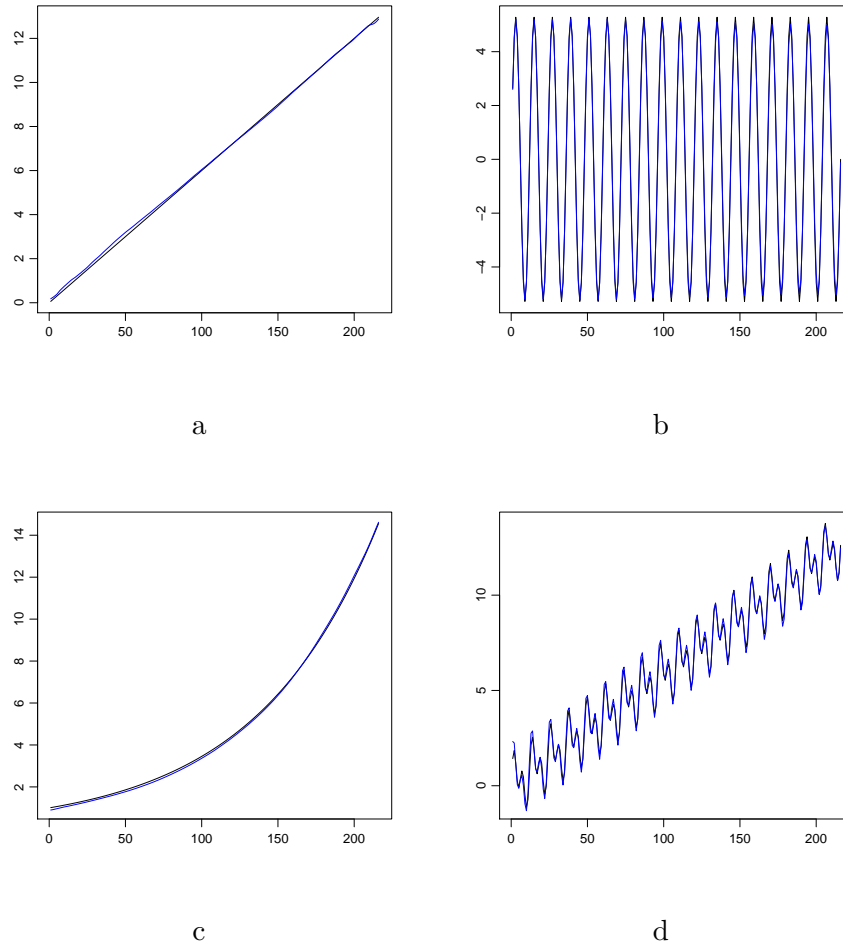


FIGURE 6.11: Original (black) and reconstructed (blue) series for cases a, b, c, and d.

to $skew(\zeta_{150})$ correspond to the signal, and the rest to the residual.

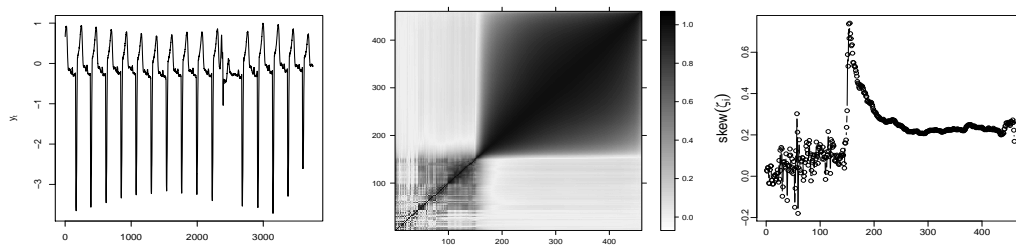


FIGURE 6.12: Noisy ECG series (left), the correlation matrix (middle) and the skewness coefficient (right) of ζ_i for the ECG series.

2. EEG:

Electroencephalography (EEG) is the recording of the brain's electrical activity. The EEG procedure is usually carried out by a highly trained specialist called a clinical neurophysiologist during a visit to a hospital. During the test, small sensors are attached to the scalp to pick up the electrical signals produced when brain cells send messages to each other. These signals are recorded by a machine and are looked at by a doctor later to see if they are unusual.

There are many different ways an EEG recording can be done. Before the test starts, the patient's scalp is cleaned and around 20 small sensors, which called electrodes, are attached using a special paste or glue. The sensors are connected by wires to an EEG recording machine. Routine EEG recordings usually take 20 to 40 minutes, although a typical appointment will last about an hour, including some preparation time at the beginning and some time at the end. Other types of EEG recording may take longer.

The main types of EEG are explained here:

- Routine EEG

A routine EEG recording lasts for about 20 to 40 minutes. During the test, the patients will be asked to rest quietly and close or open their eyes from time to time. Most of cases, they also may be required to breathe in and out deeply (known as hyperventilation) for a few minutes. At the end of the test, a flashing light can be placed nearby to see if this affects their brain activity.

- Sleep EEG or sleep-deprived EEG

A sleep EEG is carried out while the patient is asleep. This can be used if a routine EEG does not provide enough information, or to examine for sleep disorders. In some cases, the patient is required to be awake the night before the examination to aid ensure the patient will sleep while the test is carried out. It is known as a sleep-deprived EEG.

- Ambulatory EEG

This test is where brain activity is recorded during the day and night over a period of one or more days. The electrodes are attached to a small portable EEG recorder, which may be clipped on to the patient's clothes. The patient can continue with most of normal daily activities while the data is being recording.

- Video telemetry

Video telemetry, it is also called video EEG, is a special type of EEG where the patients is filmed while data is recording. This provides more details about brain activity, which is usually carried out over a few days while staying in a hospital. The EEG signal is transmitted wirelessly to a computer. The video is also recorded by the computer and kept under regular surveillance by trained staff.

EEG can help to diagnose a number of various medical conditions; for example, epilepsy, memory impairment, brain inflammation and coma. However, EEG recording is highly susceptible to different types of noise, which makes the analysis of EEG data more difficult. We use a single channel EEG that has been analysed by many authors (see, for example, [56, 251]). The signal was obtained from a set that containing 100 single channel EEG segments of 23.6-second duration. These segments were selected and cut out from continuous multichannel EEG recordings after visual inspection for artifacts, e.g., due to muscle activity or eye movements. In addition, the segments had to fulfill a stationarity criterion described in detail in [56]. The set consisted of segments taken from surface EEG recordings that were carried out on five healthy volunteers using a standardized electrode placement scheme. Volunteers were relaxed in an awake state with eyes open.

The signal was recorded with the 128-channel amplifier system, using an average common reference (omitting electrodes containing pathological activity strong eye movement artifact normal EEG segment). After 12 bit analog-to-digital conversion, the data was written continuously onto the disk of a data acquisition computer system at a sampling rate of 173.61 Hz.

Table 6.16 represents a summary of descriptive statistics for both ECG and EEG series. As can be observed from Table 6.16 that the values of mean and standard deviation (S.D) of the ECG series are smaller than their values of the EEG series. Different results can be seen for the values of the minimum and maximum of the series. It is obvious that the values of the skewness and kurtosis are also different. The results indicate that the ECG series does not have a symmetrical and normal distributions whereas the normal EEG data may does (see Figure. 6.13). This will be evaluated using different statistical tests, and will be discussed deeply in Chapter 7.

TABLE 6.16: Descriptive statistics of ECG and EEG series.

	Statistics					
	Mean	S.D	Minimum	Maximum	Skewness	Kurtosis
ECG series	-0.18	0.84	-3.72	1.00	-2.29	8.91
EEG series	6.82	42.6	-190	185	-0.18	3.54

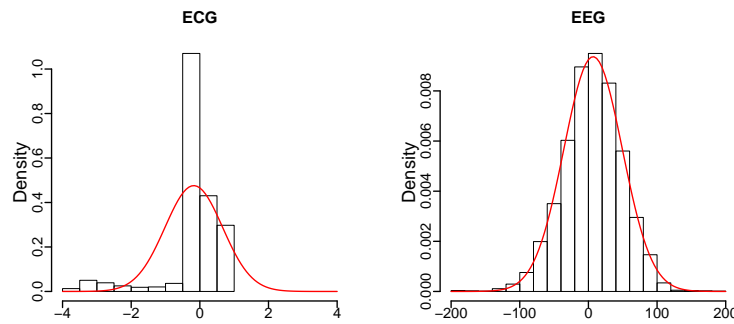


FIGURE 6.13: Histograms of the abnormal ECG (left) and normal EEG (right) series with normal curve (red line).

Figure. 6.14 (right) illustrates the EEG time series, and the results of the correlation and the skewness of ζ_i are depicted in Figure. 6.14 (middle) and (right), respectively. Both results indicate that the number of eigenvalues that correspond to the EEG signal is 91.

After finding the optimal r for each series, we can use the second stage of the algorithm. The second stage is exactly the grouping and diagonal averaging steps in SSA. Note that for grouping step, we put the eigentriples $(1 : r)$ in one group for

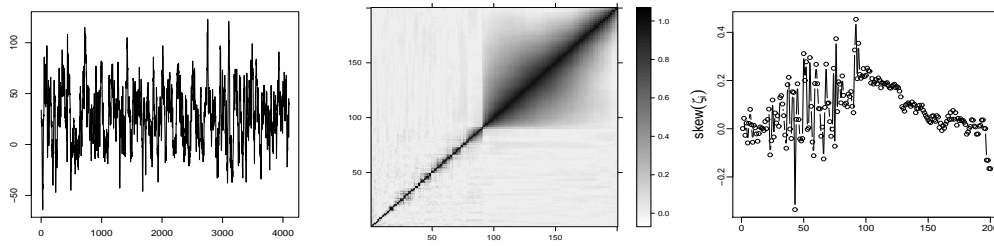


FIGURE 6.14: Noisy EEG series (left), the correlation matrix (middle) and the skewness coefficient (right) of ζ_i for the EEG series.

the reconstruction or extraction of the signal. Figure. 6.15 shows the the result of the signal extraction or reconstruction series without noise, which is obtained from eigentriples $1 : r$ ($r = 150, 91$) for the ECG and EEG series respectively. The blue and the black lines correspond to the reconstructed series and the original series respectively. Note that the first 100 observations of the ECG and EEG series were plotted in the figure for visual purposes. As a result, the selected values of r for the reconstruction of the original series are optimal.

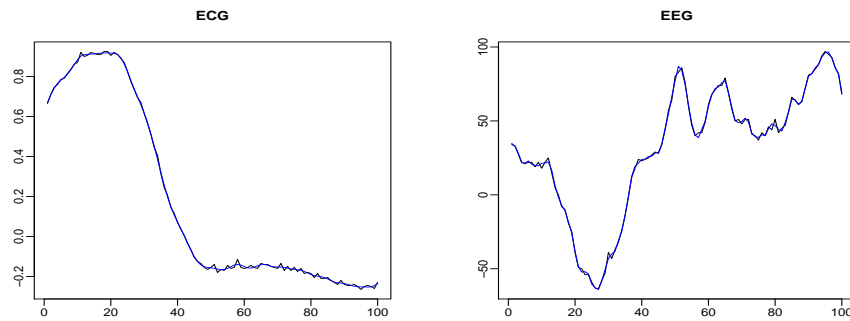


FIGURE 6.15: Original (black) and reconstructed (blue) series for ECG, and EEG series (first 100 observations).

6.4 Summary

The most important issue in singular spectrum analysis, that is, the selection of the number of eigenvalues (r) needed for reconstruction of the series, was considered. We have studied and investigated the skewness and kurtosis of the eigenvalue distribution of \mathbf{B} , and the value of variation coefficient for each eigenvalue. We have shown that the maximum value of the skewness and kurtosis coefficients for the distribution of the eigenvalues are novel indicators for selecting the optimal value of r . In addition, the correlation (or correlation matrix) between eigenvalues, ζ_i and ζ_j were considered. The correlation matrix was considered as another criterion for separability between time series components in SSA. In addition, we have used the same criteria for the choice of L . Although we have not considered all the values of L , the results indicate that the value of L should be half of the series length for best separability. The results based on these criteria have shown that the cut-off point between signal and noise components in SSA can be obtained correctly.

Chapter 7

Application in Biomedical Data

This chapter seeks to explore the applicability of the method for eigenvalue identification in four different gene expression protein profiles. In addition, the approach is applied to discrimination between normal and epileptic seizure EEG signals, extraction of strange patterns, and filtering of EEG signals and eliminating noise included in the signals.

The remainder of this chapter is structured as follows: Section [7.1](#) includes a brief introduction to noise correction in gene expression data, followed by data description and main results of the section. Section [7.2](#) presents a brief introduction to the removal of noise from EEG signals. In Section [7.2.1](#) we show that the approach can decompose the synthetic data into two main distinct subspaces. Section [7.2.2](#) presents the performance of the proposed approach in filtering EEG signals, extracting strange patterns, and discriminating between normal and epileptic EEG signals. Section [7.3](#) presents the summary of the chapter.

7.1 Noise correction in gene expression data

Segmentation in *Drosophila melanogaster* is a very important studied example of gene regulatory networks in developmental studies [[252](#)] as it has small size, short

generation time, and large brood size, which make it ideal for genetic studies. Furthermore, transparent embryos facilitate developmental studies. Model organisms are species that are studied to understand the biology of other organisms, often humans. In addition, fruit flies share 75% of the genes that cause disease with humans, so scientists can learn about human genetics by studying fruit fly genetics.

In gene regulatory networks, it is widely accepted that the pattern of the segmentation factors which has been activated by the primary morphogens direct the development of the early embryo. However, due to the presence of noise, finding the pattern of segmentation factors is not a simple task [117, 253] and even a small level of noise in gene expression patterns will considerably affect our understanding of the embryo developmental fate. Hence, it is important to probe the gene expression signal using a method which effectively enables us to filter the fluctuations of the related gene protein profile.

These profiles can mostly be achieved by using the fluorescence imaging technique [169]. Such quantification relies on the assumption that the actual protein concentrations detected by the fluorescence in situ hybridization (FISH) technique are linearly related to the embryos natural protein concentration. However, the obtained profile contains different levels of noise which need to be removed first. Among several noise removal models, SSA is a relatively new method which has recently transformed itself into a valuable tool for gene expression signal extraction. The first such application of SSA was made in 2006 when Holloway et al. studied the relation between maternal protein gradients and segmentation process in *Drosophila*, by analysing gene expression patterns extracted by SSA [164]. Two powerful characteristics of SSA are worthy of mention: there is no requirement of any assumptions about the data and related residuals, and its effective performance in noise filtering [112] makes SSA a valuable method in analysing segmentation gene profiles.

Even though the signal extraction by SSA appears to be simple, in practice it is a complicated task since in some cases the trend cannot be separated from noise or cyclic components just by choosing the first eigenvalue. This issue was raised for the first time in [117], where the author suggests the use of either a small window

length or the addition of a constant to the series to improve the separability between noise and signal [117]. Despite the practical possibility of these suggestions, there is still an open question related to the identification of the number of eigenvalues required for gene series reconstruction.

To address this question we mainly follow the algorithm proposed in Chapter 6. The proposed approach has been mainly used for noise reduction, filtering, signal extraction and distinguishing chaos from noise in previous chapters. In identifying the number of eigenvalues, this method mainly relies on the distribution of the scaled Hankel matrix eigenvalues. Here, we apply the method for signal extraction for four different genes; bicoid (*bcd*), caudal (*cad*), giant (*gt*) and even-skipped (*eve*), which are among the most important zygotic segmentation genes. The approach enables us to decide and select the appropriate number of eigenvalues related to the gene signal.

7.1.1 Real data

Data description

The gene expression data in wild-type *Drosophila melanogaster* embryos is achieved by the fluorescently tagged antibodies technique and is available via [254], where a more detailed description of the biological characteristics, method and data is given. This data was extracted from the nuclear intensities of %10 longitudinal strips and the data was not processed for any other noise removal.

Of the many segmentation genes, we are only concerned with four different ones in this study: *bcd*, *cad*, *gt* and *eve*, of which *bcd* is maternal, *cad* has both maternal and zygotic origins, and *gt* and *eve* are respectively related to gap and pair rule categories of zygotic genes [164, 169].

bcd mRNA is completely maternal and the Bcd protein gradient is formed at cleavage cycle 9 [164, 169]. Figure 7.1(a) depicts a typical example of a *Bcd* gradient

related to cleavage cycle 14(3). Although this figure suggests *Bcd* follows an exponential trend due to the high volatility seen in the series, the extraction of this trend is not a simple task.

cad mRNA has both maternal and zygotic origins and the maternal transcripts begin to translate immediately after fertilization. However, proteins encoded by *gt* and *eve* were reported to appear at cycle 12 and 10 respectively and it is accepted that the posterior stripe of *gt* expression is regulated by *bcd* and *cad* [161, 164, 169].

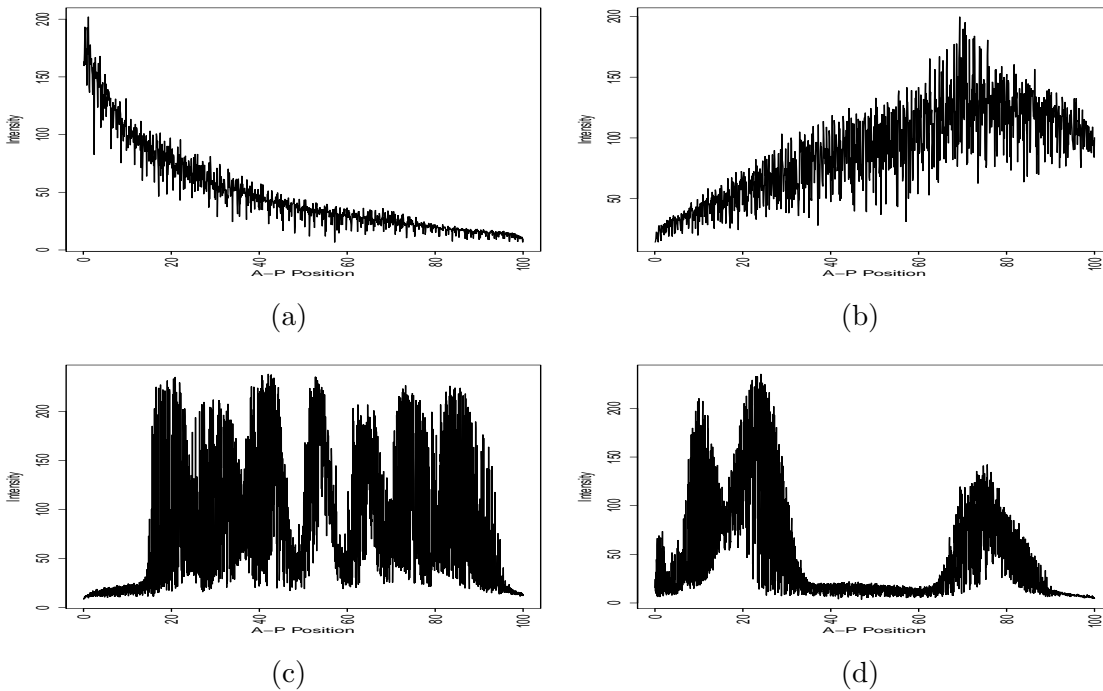


FIGURE 7.1: Experimental data from the *Drosophila melanogaster* embryo; (a): *bcd*, (b): *cad*, (c): *eve*, (d): *gt*. y axis represents gene expression profile, which tell us what actually it is doing at a point in time (x axis).

7.1.2 Main results

To generate simulated noisy profiles with similar structure, shape and distribution to the real gene expression profiles we mainly follow the algorithm presented and explained in the previous chapter. Although the gene expression profiles are slightly different from embryo to embryo, as the obtained results in terms of number of eigenvalues are similar, we only consider ten different embryos to study each gene.

In this regard, each copy of the gene expression data was simulated 10^4 times. Studying the distribution of each eigenvalue provides the capacity to obtain an accurate and deep intuitive understanding of selection of the optimal value of r . The first data for each gene is analysed and discussed in more detail, whilst the results of the other data are summarised based on the outcomes of the skewness, variation and correlation coefficients. The window length used for analysing the *bcd*, *cad*, *gt* and *eve* genes series is 200.

We mainly focus here on the *skew* and *CV* of the distribution of ζ_i . In addition, the ρ_s between the eigenvalues is also evaluated to enhance the results obtained by the *skew* and *CV* measures. The absolute value of the correlation between ζ_i and ζ_{i+1} is considered; 1 indicates that ζ_i and ζ_{i+1} have perfect positive correlation, whilst 0 shows there is no correlation between them.

Figure. 7.2 illustrates the results of *skew*(ζ_i) (left) and *CV*(ζ_i) (right) for the first data series for each gene type. It can be seen from the left column that the maximum value of *skew* is obtained for $\zeta_{c=3}$ in both *bcd* and *cad* data, whereas *skew*($\zeta_{c=4}$) is the maximum for both *eve* and *gt* series. In the right column, the results of *CV* split the eigenvalues into two groups for each data; the second group looks like a U shape which is related to the noise component. The results indicate that $r = c - 1 = 2, 2, 3, 3$ for extracting the *bcd*, *cad*, *eve* and *gt* signals, respectively.

Furthermore, the result of ρ_s can be used as a decision or test tool if the *skew* and *CV* measures give different results. However, in these typical examples, the results of those two measures are the same which also supported by the results of the correlation coefficient. It is obvious that the minimum values of ρ_s are observed between $(\zeta_2, \zeta_{c=3})$, $(\zeta_2, \zeta_{c=3})$, $(\zeta_3, \zeta_{c=4})$ and $(\zeta_3, \zeta_{c=4})$ for *bcd*, *cad*, *eve* and *gt*, respectively. Therefore, the results enhance the fact that $r = 2, 2, 3, 3$ for the first data for each gene (see Figure. 7.3).

Tables. 7.1, 7.2, 7.3, and 7.4 show the results of r based on those three measures for all 40 series. For the *bcd* signal extraction, all the outputs show $r = 2$ for all *bcd* data (see Table. 7.1). Similar results emerged in the extraction the *cad* signal, most of the outcomes indicating $r = 2$.

For the *eve* data, $r = 3$ for five series, as all the three measures give the same result. However, for example, for series 2, the results of *skew* and *CV* are different; $r = 3$ and $r = 4$, respectively. To overcome this, we look at the result of ρ_s , which confirms that $r = 4$. In this regard, the decision is that $r = 3$ for six series of ten *eve* data. Table. 7.4 demonstrates that $r = 3$ for all *gt* series except the last series, because all measures have the same results. As a result, for $L = 200$, the required eigenvalues to extract the *bcd*, *cad*, *eve*, and *gt* signals are 2, 2, 3, 3, respectively. Table. 7.5 shows the final results for all four genes along with the most frequent reported *skew*, *CV* and ρ_s .

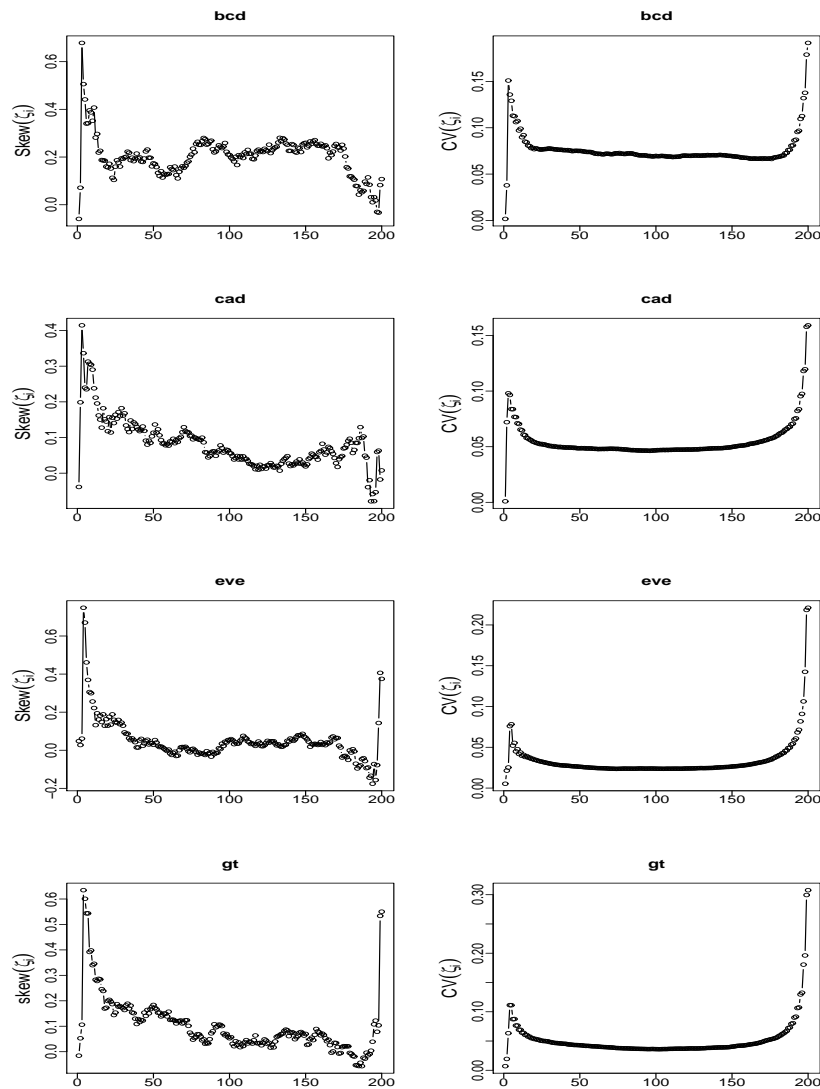


FIGURE 7.2: Skewness coefficient (left) and the variation coefficient of ζ_i (right) for the first series of *bcd*, *cad*, *eve* and *gt* data.

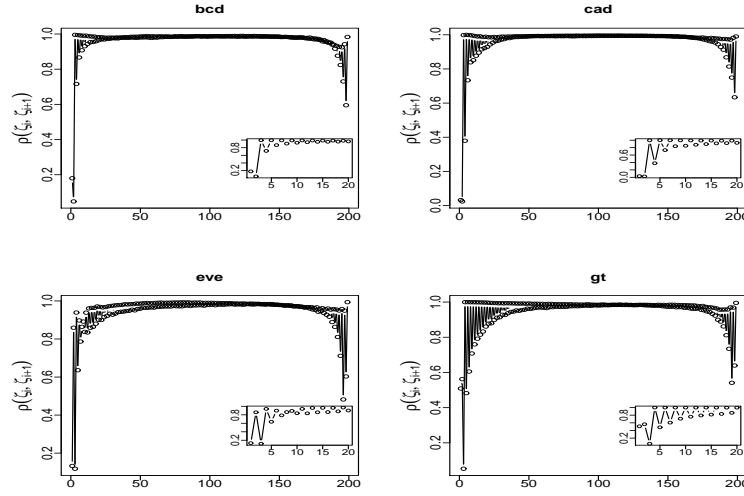


FIGURE 7.3: Correlation between ζ_i and ζ_{i+1} for the first series from each data set.

TABLE 7.1: Values of r based on *Skew* and *CV* for the ten *bcd* series.

Series	$r(skew)$	$r(CV)$	$r(\rho)$	Series	$r(skew)$	$r(CV)$	$r(\rho)$
1	2	2	2	6	2	2	2
2	2	2	2	7	2	2	2
3	2	2	2	8	2	2	2
4	2	2	2	9	2	2	2
5	2	2	2	10	4	2	2

TABLE 7.2: Values of r based on *skew* and *CV* for the ten *cad* series.

Series	$r(skew)$	$r(CV)$	$r(\rho)$	Series	$r(skew)$	$r(CV)$	$r(\rho)$
1	2	2	2	6	1	2	1
2	2	2	2	7	2	2	2
3	2	2	2	8	2	2	2
4	1	2	1	9	2	1	2
5	2	2	2	10	3	3	3

TABLE 7.3: Values of r based on *skew* and *CV* for the ten *eve* series.

Series	$r(skew)$	$r(CV)$	$r(\rho)$	Series	$r(skew)$	$r(CV)$	$r(\rho)$
1	3	3	3	6	3	4	4
2	4	3	4	7	3	3	3
3	6	6	6	8	4	4	4
4	6	4	4	9	3	3	3
5	3	3	3	10	3	3	3

After the step of identifying the value of r , we can use the leader eigenvalues in the second stage of the algorithm to reconstruct the first typical data for each

TABLE 7.4: Values of r based on $skew$ and CV for the ten gt series.

Series	$r (skew)$	$r (CV)$	$r (\rho)$	Series	$r (skew)$	$r (CV)$	$r (\rho)$
1	3	3	3	6	3	3	3
2	3	3	3	7	3	3	3
3	3	3	3	8	3	3	3
4	3	3	3	9	3	3	3
5	3	3	3	10	5	3	5

TABLE 7.5: Final result obtained in noise-signal separation study for gene data.

Gene type	$r (skew)$	$r (CV)$	$r (\rho)$
bcd	2	2	2
cad	2	2	2
eve	3	3	3
gt	3	3	3

gene. Figure. 7.4 shows the result of the gene signal extraction or reconstruction series without noise. The red and the black lines correspond to the reconstructed series and the original series respectively. As a result, the considered r for the reconstruction of the original series is obtained correctly, especially for the *bcd* and *cad* signals.

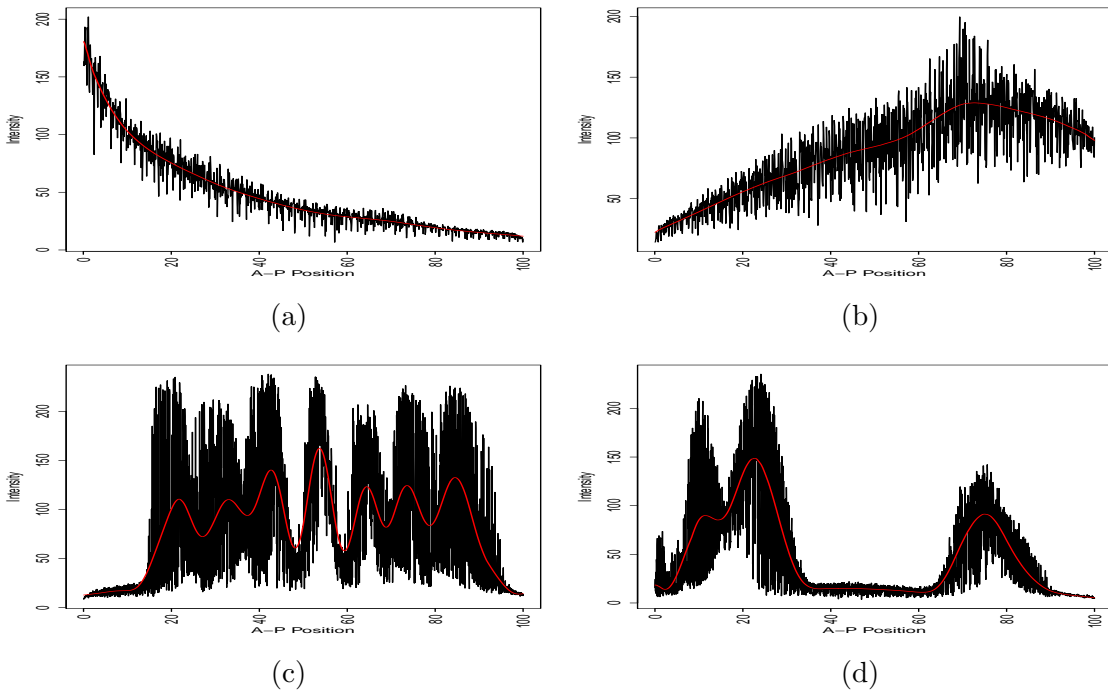


FIGURE 7.4: Original (black) and extracted signal (red); (a): *bcd*, (b): *cad*, (c): *gt*, (d): *eve* .

Taking a closer look at Figure. 7.4, it is important to note that the extracted signal profiles of *eve* and *gt* do not follow the expression data satisfactorily when the data series changes sharply. Therefore, in order to solve this issue and capture the peaks of the profiles, we used sequential SSA. The main idea underlying this approach is to apply SSA recursively on the residuals with different window length L [255]. By doing so we extract some components of the initial series using basic SSA and then extract the remaining components related to the signal by applying SSA on residuals. Such a recursive SSA application produces a gradual extraction of the signal present in the noise. Figure. 7.5 shows the result after applying sequential SSA. As can be seen, signal extraction and peak capturing have been improved accordingly.

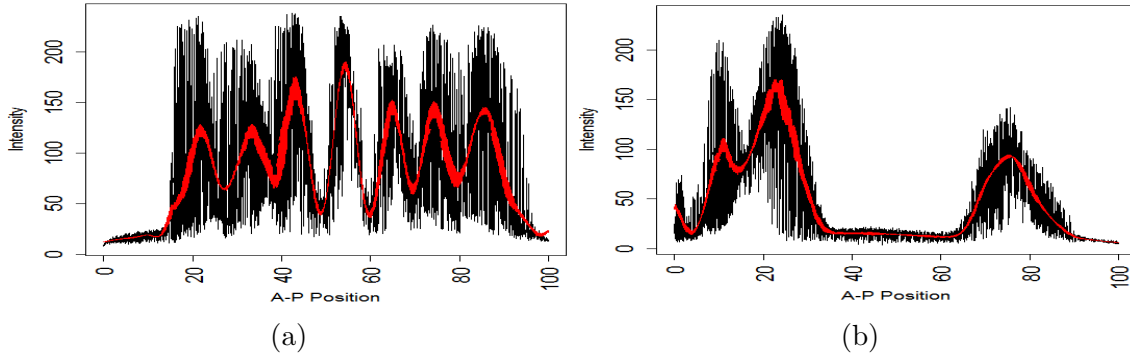


FIGURE 7.5: Improving signal extraction using sequential SSA. Original (black) and extracted signal (red);(a): *eve*, (b): *gt* .

7.2 Removing noise from EEG signal

The electroencephalography (EEG) signal is the recording of electrical brain activity, which is a complex signal, and one of the most frequently used to study and investigate neurological disorder. The EEG signals represent not only the brain function but also the situation of the whole body [256]. Furthermore, the EEG biosignal records play an important role in the detection and treatment of brain diseases, such as epilepsy and brain tumor. Moreover, the analysis of EEG can be used to diagnose brain death [257].

The EEG signal is a valuable tool to study the brain function and neurobiological disorders, however; its recording is contaminated by diverse types of noise and artifacts which can cause problems in the accurate analysis of brain signals. These types of noise can be electrical, or can be made by our bodies, since the signal records have small amplitudes [257]. Furthermore, different artifacts such as blinking of the eyes, ocular artifacts, and muscle activities make noise in EEG recording; detecting such noise becomes a complex task. Although the signals can be affected by internal and external noises, which often have unknown characteristics, the noise can be identified if the signal and noise subspaces can be accurately separated. Various methods can be implemented for denoising or removing noise from EEG signals; for instance, principal component analysis (PCA) [258]–[260], independent component

analysis (ICA) [261]–[263], and wavelet transform (WT) [264]. Furthermore, different techniques have been applied to the analysis and discrimination of different categories of EEG signals [265]–[267].

We mentioned in the first chapter that in nature many nonlinear dynamic systems can show chaotic behaviour. For example, the recording of the brain is considered as a nonlinear time series when an epileptic seizure exists, and can also be considered as a chaotic series [57]. Therefore, researchers attempt to identify strange patterns in the analysis of brain signals. In addition, analysing EEG is an important tool for detecting epileptic activity. Detection of epileptic seizure and the extracted information from brain recordings play a significant role in diagnosis and treatment.

In this subsection, we apply the approach to discriminate between normal and epileptic seizure EEG signals, to choose the optimal value of r to denoise the EEG signal segments, thereby eliminating noise, and to extract strange patterns.

7.2.1 Synthetic data analysis

Let us here illustrate the capability of the proposed approach using two other examples, as EEG recordings contain the sum of sinusoid components, noise and other signals that behave chaotically.

Example 1:

To show the applicability of the scheme, it is also used to decompose the synthetic time series generated from the well known Rossler low dimensional chaotic system:

$$\begin{cases} \frac{dx}{dt} = -y - z \\ \frac{dy}{dt} = x + ay \\ \frac{dz}{dt} = b + z(x - c). \end{cases} \quad (7.1)$$

Figure. 7.6 (left) illustrates 5000 values generated from the Rossler system, and Figure. 7.6 (right) shows the Rossler series with added white noise. In this example, the signal to noise ratio is $SNR = 14$. The idea is to discuss how the approach can

help to determine the number of eigenvalues that are required to remove the noise from the EEG series in order to perform an adequate analysis.

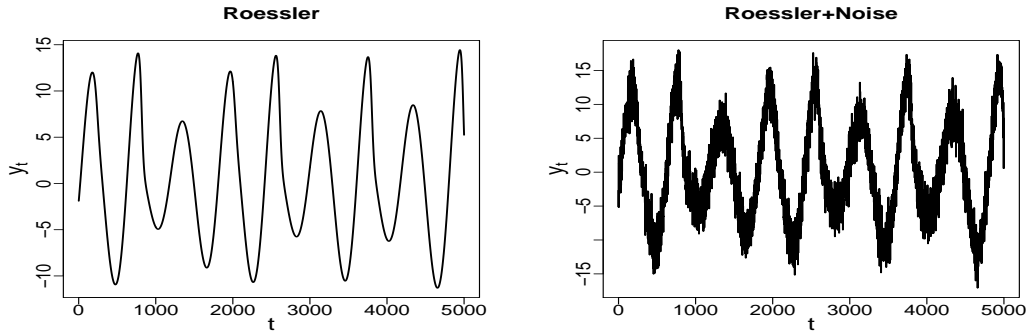


FIGURE 7.6: Rossler series (left), and Rossler + noise (right).

To apply the proposed approach, 10^5 copies of the noise component were generated and added to the signal. In this case, the independent time series Y_N^m , where $m = 1, \dots, 10^5$ has been analysed; here $L = 100$. The pattern of ζ_i ($i = 1 \dots, L$) of the matrix \mathbf{B} and its related forms for those independent time series have been studied in depth. Figure. 7.7 shows the average of the eigenvalues. It can be seen that the value of the first three are outstanding, whilst the other eigenvalues are close to each other. This is expected because the embedding dimension for a chaotic Rossler system must be at least 3. Each eigenvalue or singular value contributes to the trajectory matrix decomposition. We can consider the ratio $\bar{\zeta}_i \times 100$ as the characteristic of the matrix \mathbf{H}_i to Eq. (1.3). Thus, $100 \times \sum_{i=1}^r \bar{\zeta}_i$ is the characteristic of the best approximation of \mathbf{H} by the matrices of rank r .

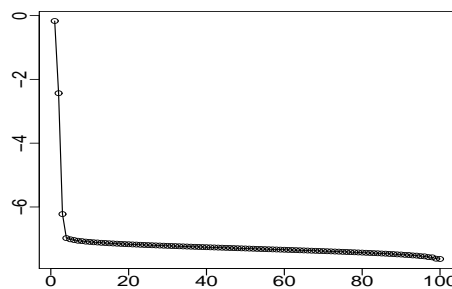


FIGURE 7.7: Logarithm of $\bar{\zeta}_i$ for the Rossler + Noise series.

Here, the corresponding eigentriples to those first three singular values can be assumed as the leading components for the original series as their ratio is 99.12. Figure. 7.8 illustrates the results of $Skew(\zeta_i)$ (top left), $Kurt(\zeta_i)$ (top right), $CV(\zeta_i)$ (bottom left), and the matrix of the absolute value of ρ_s between the eigenvalues (bottom right). The results of the coefficient of the skewness measure split the eigenvalues into two groups. It is obvious that the maximum value of $Skew$ is obtained for $\zeta_{c=4}$ and the pattern of $Skew(\zeta_4)$ to $Skew(\zeta_L)$ has the same pattern for the noise component. Therefore, the first three eigenvalues correspond to the signal and the reminder to the noise part. Similar results emerged with the other two measures, $Kurt$ and CV . It can be seen that the second group of CV results has a shape which is related to the noise (see Chapter 6). In addition, it is obvious from the correlation matrix that the first three leading eigenvalues describe the Rossler signal, and the large sparkling square is the indicator of the white noise components. Thus, the eigenvalue correlation matrix gives a clear image for the theory of the separability.

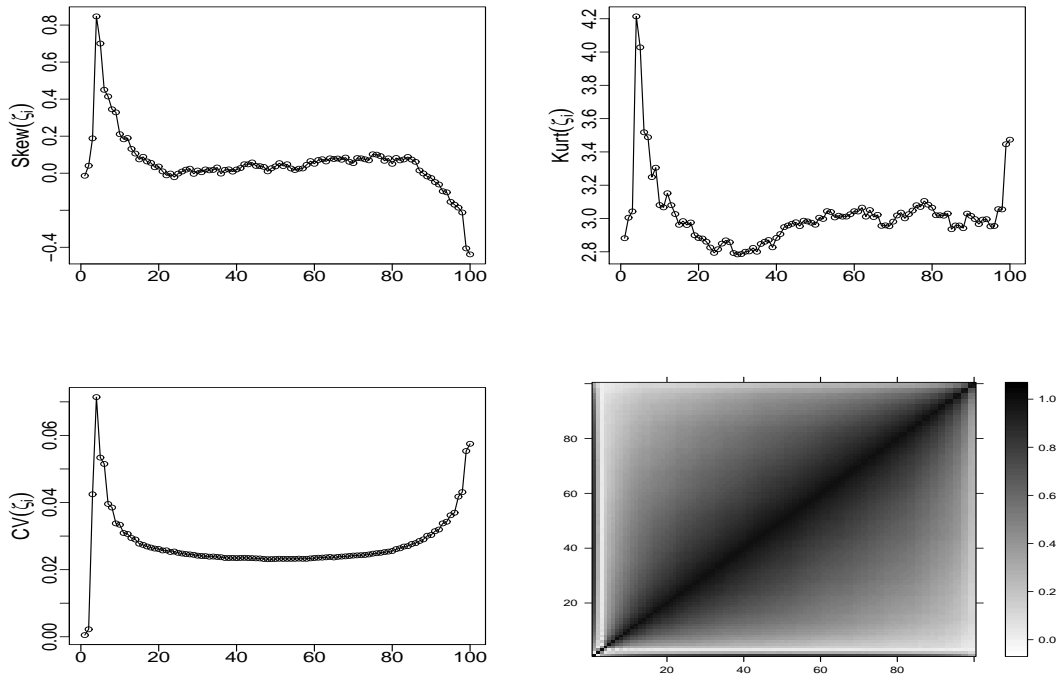


FIGURE 7.8: Results of $Skew$, $Kurt$, CV , and the correlation matrix.

The above results are evaluated using the root mean square error (RMSE) between

the original Rossler signal component and the reconstructed series by eigentriples 1– i , ($i = 1, \dots, 100$) (see Figure. 7.9 (left)). The result confirms that $r = 3$, as the minimum value was obtained between the original signal and the reconstructed series by eigentriples 1–3. Figure. 7.9 (right) depicts the reconstructed noise free series by eigentriples 1–3 (blue line), and the original Rossler component (black line).

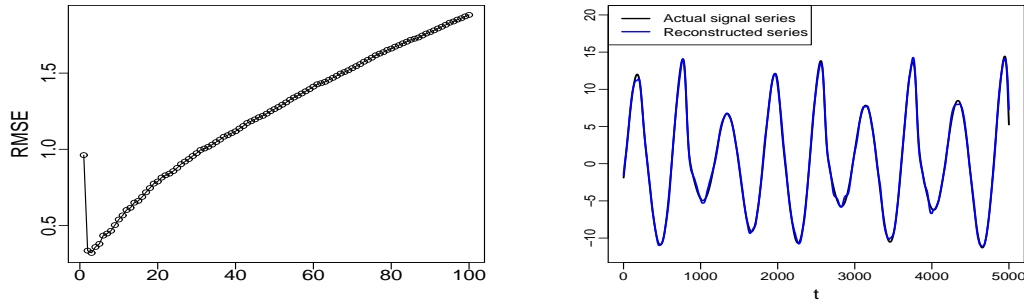


FIGURE 7.9: Root mean square error (RMSE) between the original Rossler signal and reconstructed components by eigentriples (left), and the reconstructed or extracted signal and the original signal series (right).

Example 2:

The second example is a mix of two signal components: the exponential sequence and cosine wave, with white noise:

$$y_t = s_t^{(1)} + s_t^{(2)} + \epsilon_t, \quad (7.2)$$

where $s_t^{(1)} = \exp(\alpha t)$, $s_t^{(2)} = \cos(2\pi t/T)$, $\epsilon_t \sim N(0, \sigma^2)$, $\alpha = 0.03$, $\sigma^2 = 5$, $T = 12$, and $t = 1, \dots, 96$. In this regard, the whole signal consists of the exponential trend and harmonic components. Figure. 7.10 shows a typical example of the signal $s_t^{(1)} + s_t^{(2)}$, and y_t series.

Figure. 7.11 illustrates the average of the eigenvalues. It is obvious that the value of $\bar{\zeta}_1$ is outstanding, whilst $\bar{\zeta}_2$ and $\bar{\zeta}_3$ are close to each other. This is also expected because the first eigenvalue correspond to the exponential trend, and the second and third eigenvalues correspond to the harmonic component. In this case, the

corresponding eigentriples to ζ_1 , ζ_2 and ζ_3 can be considered as the leading components for the original series as their ratio is 99.24. Figure. 7.12 show the results of $Skew(\zeta_i)$ (top left), $Kurt(\zeta_i)$ (top right), $CV(\zeta_i)$ (bottom left), and the matrix of the absolute value of ρ_s between the eigenvalues (bottom right). It can be seen that the coefficient of skewness measure separates the eigenvalues into two groups, and the maximum value of $Skew$ is observed for $\zeta_{c=4}$. Thus, ζ_i ($i = 1, 2, 3$) correspond to the signal and the reminder to the noise components. Similar results were also obtained by the $Kurt$ and CV measures. Furthermore, from the correlation matrix we can see that the first leading eigenvalue describes the exponential trend, the two pair eigenvalues correspond to the harmonic component, and the other are related to the white noise components.

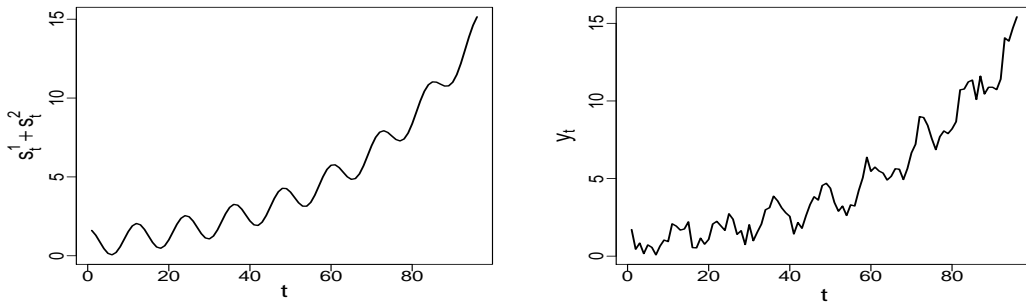


FIGURE 7.10: Typical example series, $s_t^{(1)} + s_t^{(2)}$ (left) and $y_t = s_t^{(1)} + s_t^{(2)} + \epsilon_t$ (right).

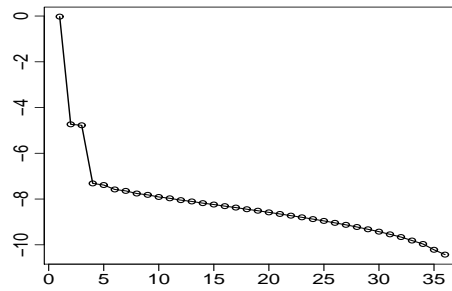


FIGURE 7.11: Logarithm of $\bar{\zeta}_i$ for the simulated series, Example 2.

The results here are also examined using RMSE between the original signal component ($s_t^{(1)} + s_t^{(2)}$) and the reconstructed series by eigentriples 1– i , ($i = 1, \dots, 36$) (see Figure. 7.13 (left)). The result of RMSE confirms that $r = 3$. Figure. 7.13

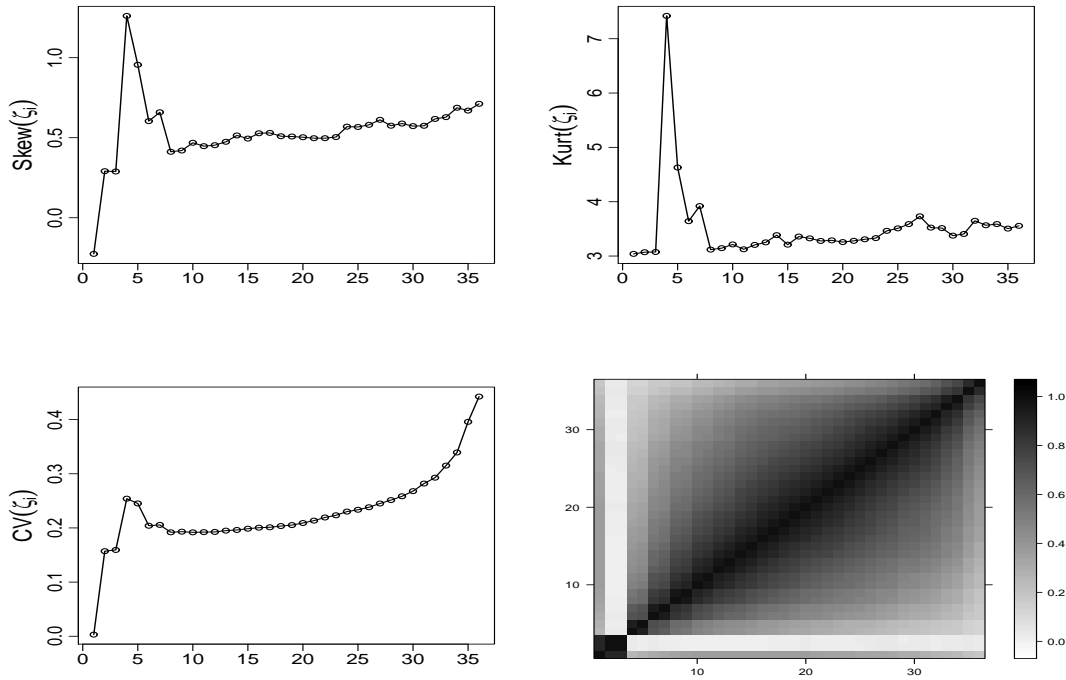


FIGURE 7.12: Results of the *Skew*, *Kurt*, *CV*, and the correlation matrix for Example 2.

(right) shows the reconstructed free noise series by eigentriples 1–3 (blue line), and the original components $s_t^{(1)} + s_t^{(2)}$ (black line). As a result, the considered r for the reconstruction of the original series is optimal, and thus the approach can be used as a promising technique for the removal noise from noisy EEG data.

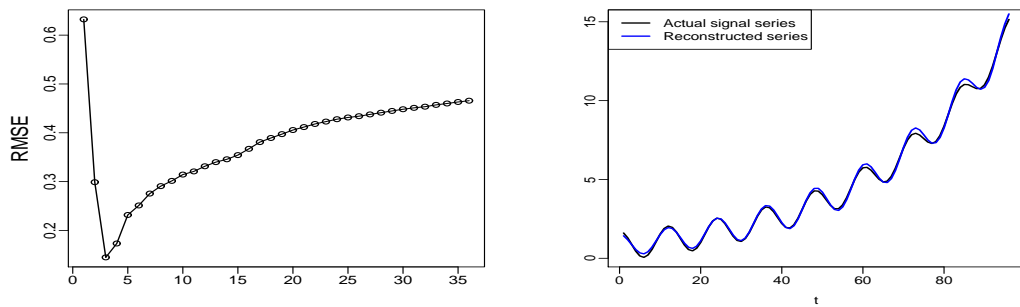


FIGURE 7.13: Root mean square error (RMSE) between the original signal and reconstructed components by eigentriples (left), and the plot of the reconstructed signal and original signal series (right), Example 2.

7.2.2 Real data analysis

7.2.2.1 EEG data selection

Two EEG signals sets [56] were used for this study. These data have been analysed by many authors as they can help us to diagnose and cure diseases in human brain, see for example [251, 268] and references therein. Each set includes 100 single-channel segments of 23.6 second duration, which were extracted from continuous multichannel EEG records after visual inspection for artifacts. The first set (normal subject) was selected from surface EEG recordings of five healthy volunteers who were relaxed in an awake state and with eyes open, using a standardised electrode placement approach. The second set contains epileptic signals (during seizure activity) from five epileptic patients.

All EEG signals were recorded with the same 128-channel amplifier system, using an average common reference (omitting electrodes containing pathological activity seizure activity segments, or strong eye movement artifacts normal EEG segments). After 12 bit analog-to-digital conversion, the data were written continuously onto the disk of a data acquisition computer system at a sampling rate of 173.61 Hz. Band-pass filter settings were 0.5340 Hz (12 dB/oct)(more details about the data can be found in [56]).

Two specimens of the normal and epileptic seizure series are shown in Figure. 7.14, and the densities of their distribution in Figure. 7.15. It is clear that the distribution of the healthy series is symmetric, whilst it is skewed for the epileptic series. It is worth mentioning that segments of the normal set can have a symmetrical distribution, whereas the distribution of the epileptic series are skewed to the right or to the left. Table 7.6 represents a summary of descriptive statistics for only series 1, series 50, and series 100 from each subject as similar results were obtained for other series. As can be observed from Table 7.6, all series of normal EEG have a smaller standard deviation (SD) than those values obtained for epileptic EEG. Similar results can also be seen for the values of the maximum and minimum of the series. It also can be seen from the table that the values of skewness (which

measures the deviation of the distribution from symmetry) is clearly different from zero for epileptic EEG series, then their distributions are asymmetrical, while those values for the normal EEG segments are almost zero, then their distributions are almost symmetrical. The D-P test is also used here to evaluate both the skewness and kurtosis of the distributions (see Table 7.7). The results based on the test confirm that all epileptic EEG segments have skewed distributions, whilst the normal EEG segments can have a symmetrical distribution, for example, series 50.

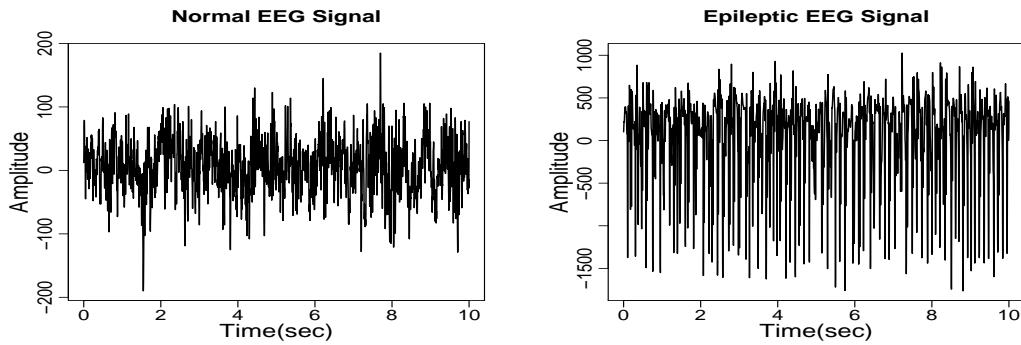


FIGURE 7.14: Specimen of the normal (left) and epileptic seizure (right) series.

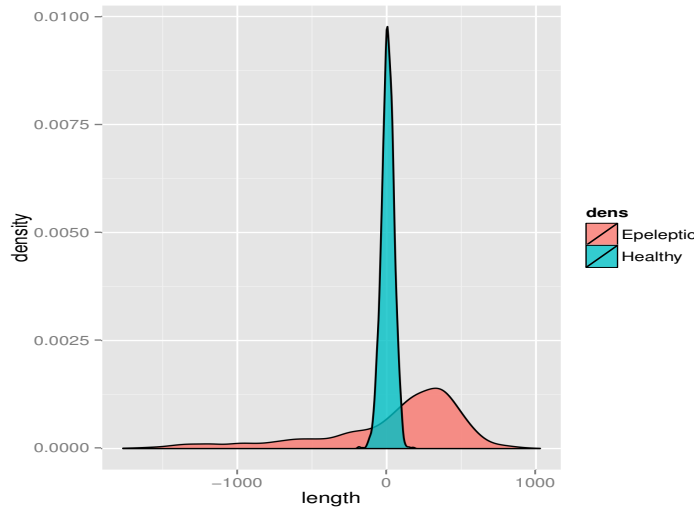


FIGURE 7.15: Density of the normal and epileptic series.

7.2.2.2 Removing noise

The main aim of the approach is the separability of the EEG signal and noise components. After selecting the optimal value of r , eigentriples that contribute to

TABLE 7.6: Descriptive statistics of for three segments; series 1, series 50 and series 100 from normal and epileptic EEG signals.

Series	Statistics					
	Mean	S.D	Minimum	Maximum	Skewness	Kurtosis
EEG.Normal 1	6.82	42.60	-190	185	-0.18	3.54
EEG.Normal 50	3.82	49.89	-206	163	-0.06	3.18
EEG.Normal 100	-28.36	42.07	-164	138	0.21	3.50
EEG.Epileptic 1	47.10	478.54	-1765	1027	-1.35	4.49
EEG. Epileptic 50	-31.14	269.93	-645	769	0.70	2.84
EEG.Epileptic 100	3.28	259.29	-833	1058	0.22	2.72

TABLE 7.7: p -value of the D-P test for three segments; series 1, series 50 and series 100 from normal and epileptic EEG signals.

Series	Statistics					
	Normal 1	Normal 50	Normal 100	Epileptic 1	Epileptic. 50	Epileptic 100
p -value	0.0018	0.337*	0.0003	<2.2e-16	<2.2e-16	0.0002

Note :*, represents symmetry based on D-P test at $p = 0.05$.

noise components are discriminated, and thus they can be separated from the signal components.

In this subsection, only two segments, from the healthy and epileptic sets, were used, simulated 10^4 times, and analysed to extract the signal or remove the noise from the EEG signal subspace. We aim here to show that the the approach can be applied for each segment separability although we will apply it for all segments together later. Figure. 7.16 (left and right columns) show the results of the measures corresponding to the normal and abnormal subjects, respectively. It is obvious that the three measures split the eigenvalues or eigentriples into two different groups (signal and noise). It can be seen that the value of $Skew(\zeta_{c=52})$ is the maximum for the non seizure signal, whilst the value of $Skew(\zeta_{c=48})$ is the maximum for the seizure signal. Furthermore, for both cases, the values of the skewness coefficient of the eigenvalues corresponding to the noise components have a slowly decreasing order. The results of the variation coefficient measure show clear evidence of the separability point between the signal and noise spaces as the values of the measure related to the noise eigenvalues have a U shape. The results of the correlation matrix between eigenvalues confirm that the number of eigenvalues corresponding to the

signal for the healthy and unhealthy subjects are 51 and 47, respectively. Thus, by using the eigentriples corresponding to those eigenvalues we can reconstruct the signal, and remove the noise part. Figure. 7.17 shows the extracted signals, and the noise series after extracting the signal from the original series for both subjects. We should mention that the same procedure can be done for other segments. Similar results were obtained by considering each segment separately.

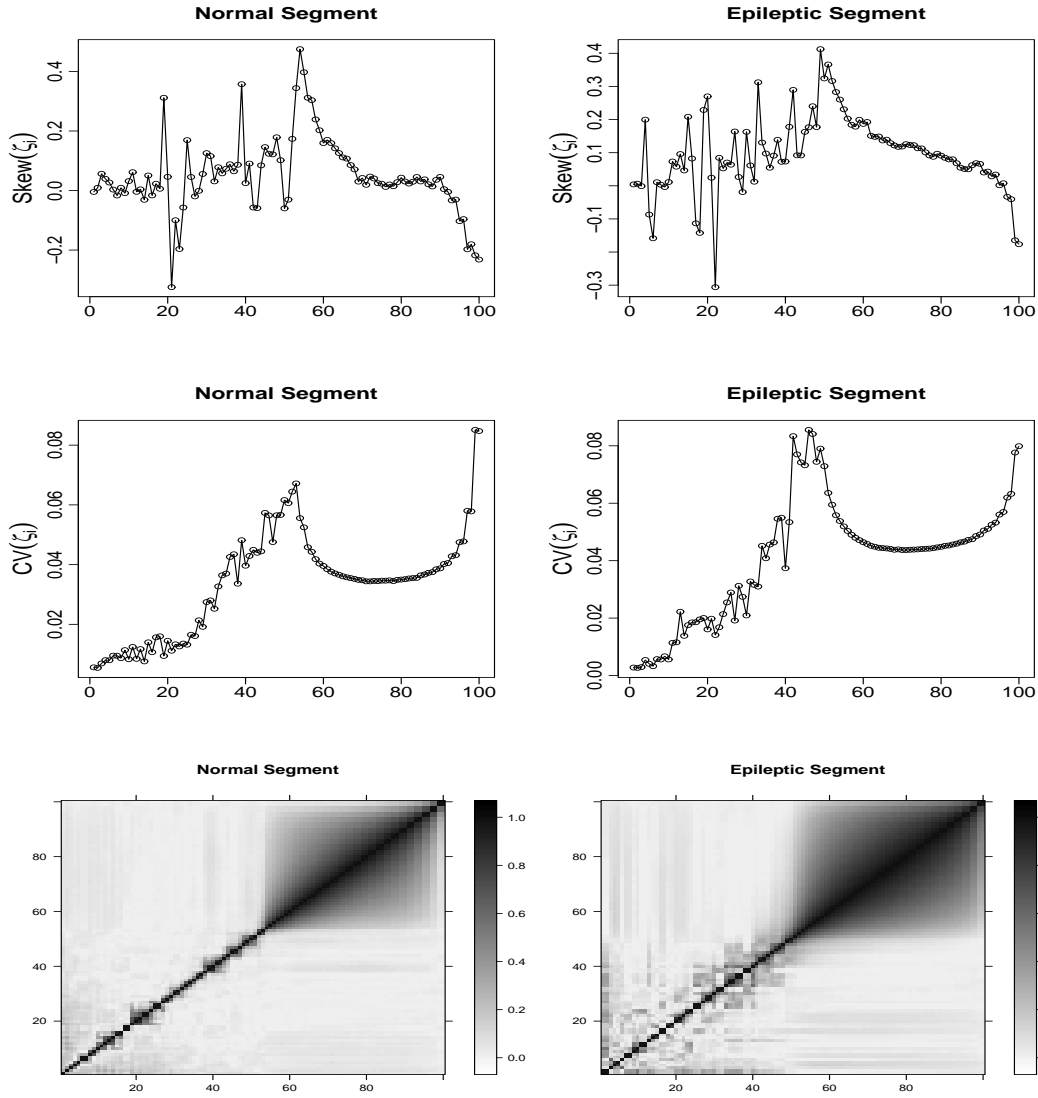


FIGURE 7.16: Results of all measures corresponding to the normal and abnormal EEG series.

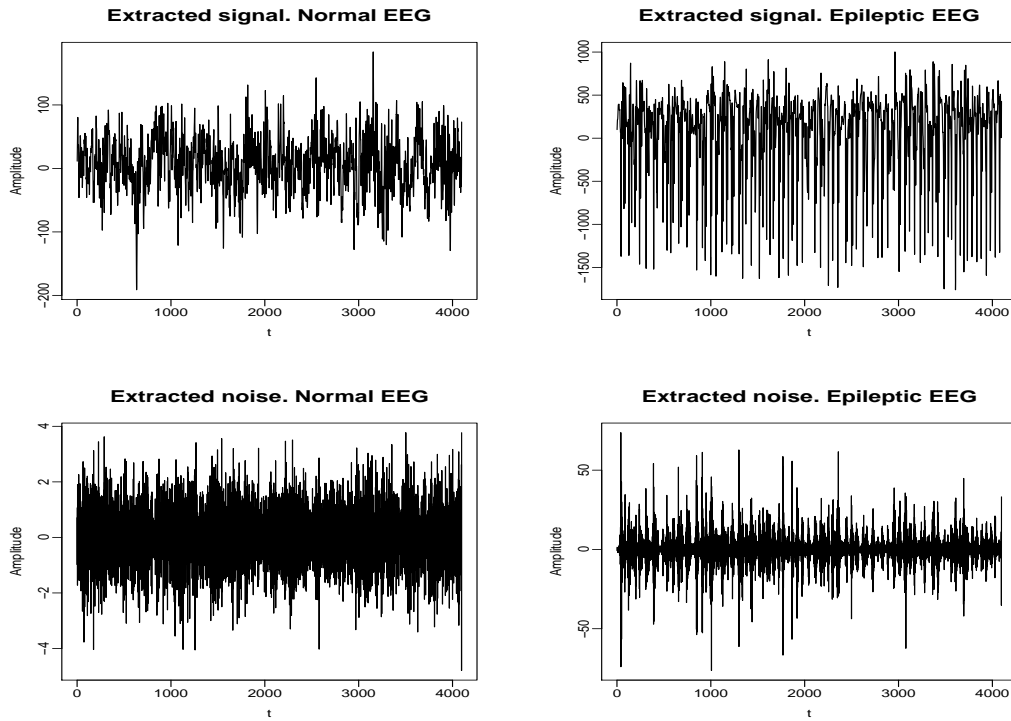


FIGURE 7.17: Extracted signals (top) and extracted noise (bottom) for healthy and epileptic seizure EEG.

7.2.2.3 Strange patterns

The main and challenging question in the nervous system is how one can differentiate between deterministic and stochastic patterns. In Chapter 5, we showed that the distribution of the scaled Hankel matrix eigenvalues can be used as a novel approach for distinguishing chaos from noise. It has been applied for discriminating between Lorenz, Henon, Tent, and white noise. The applicability of the approach was evaluated using ECG time series. The results of the approach confirmed that there is evidence of chaotic behaviour in that series, and we found an attractive pattern using the embedding method. There are several research studies devoted to the analysis of experimental data by means of the tools of nonlinear dynamics that are related to the embedding method and the delayed embedding which we use in the present study (refer to [184, 185]).

In this subsection, we only have used the first component from the extracted signal components, and the residual series for both segments, which were determined in

the the previous subsection, to detect strange patterns or chaotic behaviour in EEG signals. The two reconstructed series by the first eigentrple, and two residual series are used, and then the delayed embedding method is applied . In Figure. 7.18 (top), it can be seen that the scatter plot of y_t with time delay 10 has a very nice attractor for both reconstructed signals, which are different from each other. It can also be observed that the orbits produced move in different ways. In the healthy case, the orbit moves between the values -20 and 20, but between -300 and 300 in the unhealthy case. By looking at the scatter plot of residual series with time delay 20 for both subjects, we can see a clear pattern only for the residual series obtained from the seizure signal (see Figure. 7.18 (bottom)).

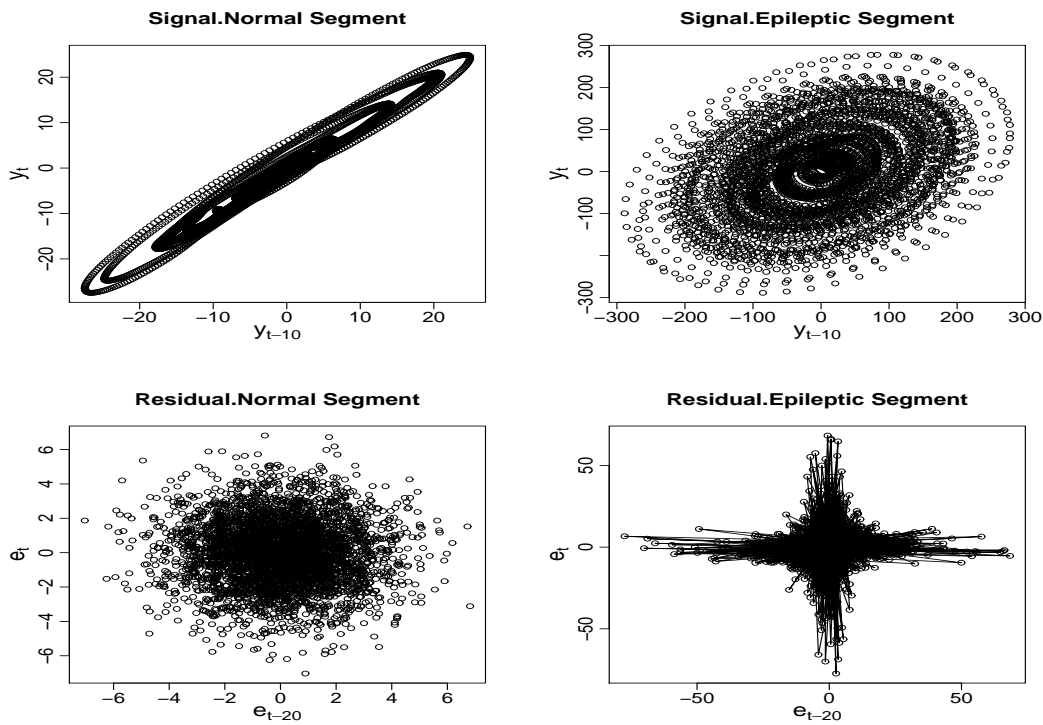


FIGURE 7.18: Plot of y_t with time delay 10 for both reconstructed series, and with time delay 20 for the two residual series.

By looking at Figure. 7.18 (right bottom), we can see that the extracted noise of epileptic EEG signal may have some remaining signal components. In order to solve this issue and extract those remaining components, we applied SSA again with the same window length. The main idea underlying this technique is to apply SSA recursively on the residuals [255]. Thus, we can extract some components of the initial series using basic SSA and then extract the remaining components related to

the signal by applying SSA on residuals. Such a recursive SSA application produces a gradual extraction of the signal present in the noise. Figure 7.19 shows the result after applying SSA again on the residual. It can now be seen that the extraction of the signal and noise components have been improved accordingly.

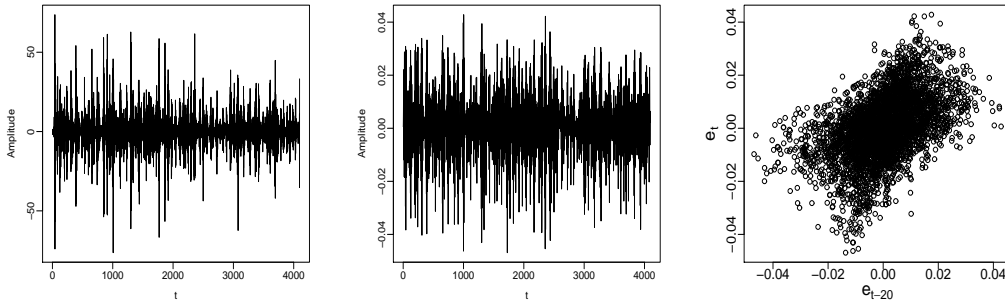


FIGURE 7.19: Improving signal (left) and noise (middle) extractions of epileptic EEG signal, and plot of y_t with time delay 20 for the noise series (right).

It should be mentioned here that we can distinguish between EEG signals by looking at the explored strange patterns, and studying the distributions of the reconstructed and residual series as shown above. Although it is possible to simulate and analyse each segment and use all the filtered EEG segments and residuals to discriminate between EEG classes and detect epileptic seizure, we can differentiate between EEG subjects by applying the proposed approach to the EEG segments directly without using a simulation method for each segment, and using different criteria for the discrimination of them. This will be discussed in the following subsection.

7.2.2.4 Discrimination of EEG signals

This subsection demonstrates the ability of the approach to distinguish between EEG signals by using all 100 segments from each class. Each single channel from each class is analysed and ζ_i is calculated. Note here that we do not need a simulation technique as each class has 100 single channels, which can be considered as simulated series, and are sufficient for this study. Consequently, for each class the average of each eigenvalue of matrix \mathbf{B} can be obtained, and other statistical tests can be applied to each eigenvalue data. The same measures that have been used

in the previous section are applied here. They are used as features to recognize the difference between normal and seizure EEG signals. The potentiality of these measures to differentiate between the two signals depends on the values of those measures. The results of these measures are similar to those shown in Figure. 7.16. The results indicate that the maximum value of $Skew$ is obtained with ζ_{52} for the normal subject whilst with ζ_{48} for the epileptic one. It is clearly observed that there is a significant difference between the values of $Skew$ for the normal and epileptic conditions, especially, between the first two values. Similar results emerged for both subjects using the kurtosis measure. Furthermore, the first values of $Kurt(\zeta_i)$ are similar for both cases, whereas there are big differences between the last 50 values (see Figure. 7.20 (middle)). In addition, by looking at the results of CV , especially the last 10 values, we can see that CV increases for the epileptic signal whilst it decreases for the normal signal.

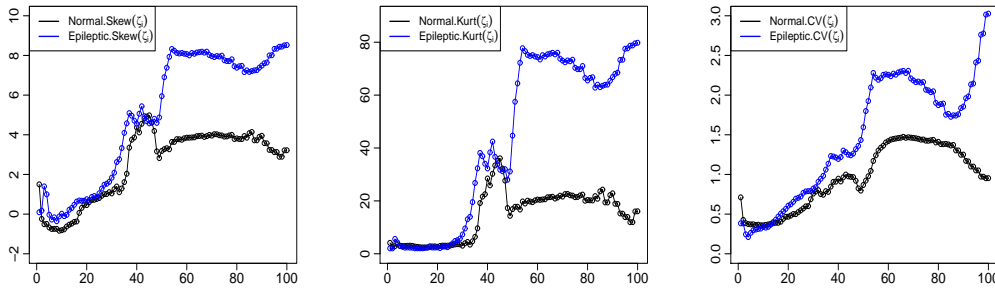


FIGURE 7.20: Results of $Skew(\zeta_i)$, $Kurt(\zeta_i)$, and $CV(\zeta_i)$ for normal and epileptic EEG segments.

Let us now consider only the largest eigenvalue obtained from the EEG segments for both subjects. Let $\zeta_{1,he} = \{\zeta_{1,n}^{he}\}_{n=1}^{100}$ and $\zeta_{1,ep} = \{\zeta_{1,n}^{ep}\}_{n=1}^{100}$ denote the largest eigenvalue for the 100 normal and epileptic segments, respectively. The distribution of $\zeta_{1,he}$ and $\zeta_{1,ep}$ are different, $\zeta_{1,he}$ has approximate skewed distribution, whereas $\zeta_{1,ep}$ has a symmetric distribution (see Figure. 7.21). It is interesting here that the distribution of ζ_1 for the healthy subject has a fat-tail. To evaluate this, two statistical tests are also applied to the distribution of $\zeta_{1,he}$ and $\zeta_{1,ep}$: D-P and K-S. As stated earlier, the D-P test is used to evaluate both the skewness and kurtosis

of the distribution. The K-S test can be used to compare a sample with a known probability distribution, or to compare two samples.

Table 7.8 provides information about the p -values of the D-P test for $\zeta_{1,he}$ and $\zeta_{1,ep}$. It is obvious that the p -value is significant for the healthy case, 0.001 for $\zeta_{1,he}$, whilst 0.788 for $\zeta_{1,ep}$. Thus, the null hypothesis is accepted that the data of $\zeta_{1,ep}$ (epileptic seizure set) is symmetric, whereas it is rejected for the data of $\zeta_{1,he}$ (normal set), and as a result the distribution of $\zeta_{1,he}$ is skewed. In addition, the p -values of the K-S test is significant; p -value=0.0014, which indicates that the data of $\zeta_{1,ep}$ and $\zeta_{1,he}$ come from different distributions, and thus their corresponding signals are distinguishable. Similar results were obtained between $\zeta_{i,he}$ and $\zeta_{i,ep}$ ($i = 50, 100$), (see Table 7.9). Similar results were also found for other eigenvalues, thus they are not reported here.

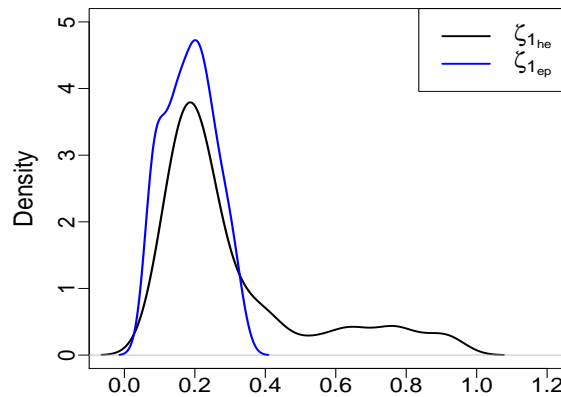


FIGURE 7.21: Density of of the distribution of $\zeta_{1,he}$ (normal) and $\zeta_{1,ep}$ (epileptic).

TABLE 7.8: p -value of the D-P test for $\zeta_{1,he}$ and $\zeta_{1,ep}$.

	p -value
$\zeta_{1,he}$	0.001
$\zeta_{1,ep}$	0.788*

Note :*, represents symmetry based on D-P test at $p = 0.05$.

TABLE 7.9: The p -value of KS test between ζ_i ($i = 1, 50, 100$) for the normal and epileptic EEG series.

	The p -value of KS test		
	$\zeta_{1,he}$	$\zeta_{50,he}$	$\zeta_{100,he}$
$\zeta_{1,ep}$	0.0014		
$\zeta_{50,ep}$		0.004	
$\zeta_{100,ep}$			0.001

Finally, the maximum, minimum, mean, and standard deviation were used as features to also distinguish between these two sets. Table. 7.10 shows the extracted features of 200 records from the two classes. It can be seen that the results are totally different from each other. Thus, these features can be used as useful information in classifying the EEG signals.

Remember that here we have considered all segments together in each case, however, we can use each segment from each case, simulate it and find the results. The results are similar to what found in Table 7.10, which shows those used features are totally different for the two subjects, and thus their corresponding signals are distinguishable.

TABLE 7.10: Extracted features of 200 segments from the two classes.

Subjects	Extracted features	Largest eigenvalue of \mathbf{B}
Healthy	Maximum	0.9237
	Minimum	0.0884
	Mean	0.3066
	Standard deviation	0.2181
	Coefficient of Variation	0.7114
Epileptic Seizure	Maximum	0.3318
	Minimum	0.0629
	Mean	0.1859
	Standard deviation	0.0713
	Coefficient of Variation	0.3837

7.3 Summary

To show the applicability of the scheme, we have also used it here to analyse synthetic time series generated from the well known Rossler low dimensional chaotic system, and cosine wave. The results of simulated data were evaluated using a statistical test between the original Rossler signal component and the reconstructed series by eigentriples.

The approach was examined in using real data. It was applied to remove the noise and signal extraction in four different *Drosophila* segmentation genes. We have demonstrated that the approach gives promising results for extracting the signal of gene expression and also indicating that the method used for removing noise from the protein profile of gene expression should be flexible enough for different type of genes, as the results have shown that we need a different number of eigenvalues for signal extraction in each gene.

In addition, the approach was examined in distinguishing between epileptic seizure and normal EEG series, and denoising the EEG signals. Two EEG signal sets were used, each set including 100 single-channel segments. We used different criteria to choose the appropriate value of r for separability between the noise component and EEG signal, extracting strange patterns and discriminating between normal and epileptic EEG.

First, only two specimens from the normal and epileptic seizure sets were studied to determine the value of r and remove noise from the signals. After finding the value of r corresponding to the signals, we applied the embedding method to the first component from the extracted signals, and the residual series in both subjects, to detect strange patterns. It was found that the scatter plot of y_t with a time delay has a nice attractor for both reconstructed signals which are different from each other and that the produced orbits from both signals move in different ways. It was found that there is a strange pattern only for the residual series obtained from the epileptic seizure signal, which indicates the possibility of chaotic behaviour in the epileptic signal.

Furthermore, we have considered all segments from each subject. It was observed that there is a difference between the values of *Skew* for the normal and epileptic conditions, especially, between the first two values. Similar results emerged for both subjects using the kurtosis measure. For both cases, the first values of *Kurt* are similar whilst there are big differences between the last values. Moreover, by looking at the results of *CV*, particularly the last values, it was observed that *CV* increases for the epileptic signal, whilst it decreases for the normal signal.

In addition, various criteria based on the largest eigenvalue were also used as new features to distinguish between the two signals. It was found that the results of the two signals are totally different from each other. Therefore, these features can be considered as useful information to classify brain signals.

Chapter 8

Conclusions and Future Research

Here we draw conclusions from the main findings of the thesis, followed by our ideas for future research; theoretical and experimental work.

8.1 Conclusions

Singular spectrum analysis is considered as a promising and reliable technique of time series analysis and forecasting. The main concept in studying SSA properties is separability, which characterises how well different components can be separated from each other. Separability depends on the selection of L and r , which is the main issue in the use of SSA; an improper choice of L and r would imply inferior separability. As our interest is to consider the signal as a whole, the importance of the selection of L becomes less important. Thus, the main aim of the thesis was to introduce a new approach for choosing the value of r , which relies on the distribution of the scaled Hankel matrix eigenvalues. The selection of L was then considered. This makes a novel contribution to the area of noise reduction and filtering in biomedical signal processing and time series analysis, as biomedical signals are often corrupted by artifacts and noise, which require separation or signal extraction before any statistical and medical evaluation. Another challenge in analysing biomedical signals is that their data is often non-stationary, particularly when there is an

abnormal event observed within the signal, such as epileptic seizure, and they can also present chaotic behaviour. Distinguishing chaos from noise is one of the main issues in the modern age because they share common properties, which in turn make them indistinguishable. Therefore, another important aim of the thesis was to provide a viable solution to this problem by the use of the distribution of the scaled Hankel matrix eigenvalues for the differentiation between and identification of chaos from noise.

In this research, we have presented a new approach and some theoretical properties of the eigenvalues of a scaled Hankel matrix that can be used for identifying signal subspace from a noisy time series. First, we presented the theoretical aspects of singular spectrum analysis and some mathematical background. We introduced and included new theoretical results and proposals on the eigenvalues ζ_i of matrix **B**. Several properties were proposed and proved, which can help us to distinguish between signal and noise components. The theoretical results indicated that the distribution of the eigenvalues can help us to distinguish between signal and noise, then to identify the eigenvalues corresponding to the signal subspace for signal extraction. The theoretical results also included the effect of the window length and rank of the Hankel matrix on its eigenvalues. We considered a number of cases with different values of L and rank of **H** to explore the relationship between eigenvalues. The results indicated that the largest and smallest eigenvalues have respectively a decreasing and increasing order for $2 \leq L \leq (N + 1)/2$, whilst having respectively increasing and decreasing orders for $(N + 1)/2 \leq L \leq N - 1$. In addition, the change and increase in window length affect and change the distribution of the smallest eigenvalue, but do not affect the distribution of the largest one for the white noise process.

Second, the simulation study demonstrated that the approach performs very well in the comparison between signal and noise series based on the distribution of the eigenvalues. The distribution of the eigenvalues of matrix **B** generated from different distributions was studied, and several properties were introduced. We have considered various cases: symmetric and nonsymmetric distributions, trend series and sine waves. The results indicate that for a large length of series N ,

the average of each eigenvalue converges to $1/L$ for a white noise process, whereas this result has not been found for other cases. Furthermore, the variance of the largest eigenvalue is approximately equal to the variance of the smallest one and the pattern of variance for all eigenvalues has a U shape for the white noise process. In contrast, various results were obtained for the other cases. The results also confirm that for the white noise process the distribution of the first eigenvalue is skewed to the right, whilst it can be symmetric for the trend and nonsymmetric distributions. The results also confirm that, for all cases, the distribution of the middle part of ζ_i can be symmetric, whilst the distribution of the last eigenvalue can be skewed. It was found that for small window length, the smallest eigenvalue has negative skewed distribution for all cases, except the trend and sine ones. In addition, the correlation between eigenvalues was assessed. It was found that the correlation matrix between ζ_i and ζ_j for a white noise process is different to what was observed for other cases, in that they had a complex structure.

Furthermore, the effect of the parameter of normal distribution μ and σ^2 on the eigenvalues of \mathbf{B} were evaluated. Based on the results, the distribution of ζ_i for a white noise process, $N(0, \sigma^2)$, is always skewed to the right, which is totally different from the one obtained by the sample generated from $N(\mu, \sigma^2)$, $\mu \neq 0$. The parameters of normal distribution or the rate σ^2/μ can play a significant role in determining the symmetry of the distribution of each eigenvalue. In addition, the distribution of the smallest eigenvalue for the white noise process converges from left to right with the increase in window length, which can also be considered as a feature to determine the noise part in a noisy time series, thereby making it possible to extract the desired signal.

Third, the distribution of the eigenvalues was used as a novel approach to the distinction between deterministic chaotic systems and stochastic processes. Using the proposed approach we were able to distinguish between chaotic series, Henon, Logistic, Lorenz and Tent, and white noise series. It has been indicated that the distribution of ζ_i ($i = 1, 2$) is symmetric for chaotic cases, particularly Henon and Logistic, whilst skewed for the white noise, Lorenz and Tent series. The distribution of ζ_i for the white noise is totally different to what obtained for the Tent map, as

the direction of the skewness is different. Although the distribution of ζ_1 and ζ_2 are skewed for the Lorenz series, which are similar to their distributions for the white noise process, it is possible to determine whether the process is deterministic or stochastic by calculating the coefficient of skewness or by looking at the matrix of the correlation. It was observed that the correlation matrix between the eigenvalues for the white noise series was totally different to what emerged for the chaotic series. In addition, different statistical test were used to to verify the results. The results confirm that the distribution of eigenvalues for the WN and chaotic series do not share the same distribution. The application in real time series also confirmed the validity and applicability of the proposed method to answer the question of whether chaos and noise are distinguishable.

Fourth, we have developed the approach and used the skewness and kurtosis coefficients of the eigenvalue distribution, together with the coefficient of variation and the matrix of correlation between eigenvalues as new indicators and criteria for the separability between signal and noise components. Based on the results of these criteria, we were able to choose the value of r for separability between the main components, signal and noise. Thus, the eigentriples corresponding to the r eigenvalues can be used to reconstruct the time series. In addition, based on the approach and its criteria the value of the window length should be large and can be half of the series. The results obtained by the introduced criteria coincided with the results obtained by other criteria used in the thesis. As a result, the approach and its criteria can help us to answer the question of what the optimal values of L and r in SSA are.

Finally, to also show the applicability of the scheme and its successful applications, it was used to analyse real biomedical data. It was applied to remove noise and to extract signals in different *Drosophila* segmentation genes. The criteria were used to identify the eigenvalues corresponding to gene signal from those corresponding to the noise. The results confirm that the proposed approach gives a promising output for the gene expression signal extraction and also indicates that the method used for removing noise from the protein profile of gene expression should be flexible enough

for different types of genes, as in this study we have obtained different numbers of eigenvalues needed for signal extraction in each gene.

In addition, we presented another application of the approach. This includes the analysis of EEG signals, noise and artifact reductions, discrimination between EEG signals, and extraction of strange patterns. The approach was examined in discriminating between normal and epileptic EEG segments, and filtering the EEG signals. Two sets from normal and epileptic EEG signals were considered, each set containing 100 single-channel segments. In the first stage, only one specimen from each set was analysed to identify the value of r and remove noise from the signals. After obtaining the value of r and extracting the signals, the embedding method was applied to the first component from the extracted signals, and also to the residual series for the detection of strange patterns or chaotic behaviour. The results indicated that there are attractive patterns for both reconstructed signals which are totally different from each other, but there is only a clear strange pattern for the residual series obtained from the epileptic seizure signal, which means the possibility of chaotic behaviour in the epileptic EEG signal.

In the second stage, all the segments from each subject were then used. Various measures were used to recognize the difference between normal and seizure EEG signals. The potentiality of these measures to differentiate between the two signals depends on their values. The results clearly indicate that there is a significant difference between the values of the measures used for the normal and epileptic subjects, which can be an answer to the question of whether we can differentiate between normal and abnormal biomedical signals.

In addition to all the above results, we introduced several criteria based on the largest eigenvalue, and used them as novel features to classify EEG signals. The results indicated that the two signals are distinguishable. Thus, the recommended features can also be used as useful information in classifying brain signals.

As a result, all our findings confirm the impressive performance of the proposed approach in distinguishing between chaos and noise and selecting the value of r for identifying the signal subspace from a noisy time series. The approach does

not need any assumptions; for example, stationarity of the series or linearity of the residual. This shows that the proposed approach can be considered as a promising one for the extraction of any biomedical signal contaminated in linear, nonlinear, stationary, or non-stationary noise, or buried in other signals. Furthermore, the main findings confirm that the proposed approach can be interesting to other researchers in different areas and disciplines where separability, distinction between noise and signals, noise reduction and filtering matters.

8.2 Future Research

For future research, the theoretical distribution of matrix \mathbf{B} is of interest. Therefore, we aim to compare the distribution of matrix \mathbf{B} with several well known theoretical distributions. Furthermore, the theoretical distribution of the smallest eigenvalue is also of great interest, because, for instance, its behaviour is used to prove its convergence to the circular law. Accordingly, study of the local properties of the spectrum, as well as the related distribution, is of interest.

Furthermore, we believe that by using the proposed approach it is also possible to find the optimal value of the window length in singular spectrum analysis to analyse and forecast various time series. Therefore, we aim to evaluate the applicability of the results found here for noise reduction of other real time series that display chaotic behaviour in different disciplines. Moreover, the presence of outliers is an important issue in time series analysis; they make the time series non-stationary and could mislead forecasting results. We thus aim to evaluate the approach in the existence of outliers, where Gaussian distribution is considered as the noise distribution.

Additionally, we will apply the properties obtained here as extra criteria for filtering series with more complex structures and for considering different types of noise; for example, red noise. We may also design a statistical test based on the result found of the determinant of matrix \mathbf{B} to examine the separability obtained by the approach. Furthermore, we may use other decomposition methods such as tensor

decomposition, instead of singular value decomposition, to examine and compare their performance with the approach.

Appendix A

Linear and Non-Linear Dependencies

The aim of using linear and nonlinear measures is to test if there is a linear or non-linear relationship between eigenvalues, and to explore if there are any similarities and differences when using these measures.

A.1 Linear correlation coefficient

A linear measure of dependencies such as linear correlation can be used to measure dependencies between two time series in linear dynamic systems. The three coefficient linear correlations that are commonly used to test dependency between two random variables are Pearson, Spearman and Kendall.

A.1.1 Pearson correlation

One of the most famous linear correlation measures is the Pearson correlation coefficient. This coefficient, ρ , between two random variables X and Y is defined as:

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E(X - \mu_X)(Y - \mu_Y)}{\sigma_X \sigma_Y}. \quad (\text{A.1})$$

where E is the expected value operator, μ_X , σ_X , and μ_Y , σ_Y are expected values and the standard deviation of random variables X and Y , respectively. The sample linear correlation coefficient of N observations of random variables X and Y can be obtained by replacing μ_X and μ_Y with the sample means \bar{x} and \bar{y} and also σ_X and σ_Y with the sample standard deviations s_x and s_y (as estimators of μ_X and μ_Y in Eq. A.1, respectively). However, observed data might not be normally distributed and might tend to have marginal distributions with heavier tails. Pearson's correlation coefficient, ρ , can only capture linear dependencies, and can not detect even a simple nonlinear association. Moreover, it has a large number of disadvantages; see [25].

A.1.2 Spearman correlation

The Spearman rank correlation, ρ_s , is similar to that of Pearson, and measures the linear relationships between ranked variables rather than between their raw numbers. It is a nonparametric measure of the strength of the linear association between two scale variables, and can be applied to ordinal data as well. It can measure how consistently one variable increases or decreases as a second variable increases (monotonic). For a sample of size N , the N raw scores X_i , Y_i are converted to ranks x_i , y_i , and ρ_s is computed as follows:

$$\rho_s = 1 - \frac{6 \sum d_i^2}{N(N^2 - 1)} \quad (\text{A.2})$$

where $d_i = x_i - y_i$, is the difference between ranks.

A.1.3 Kendall correlation

The Kendall correlation, τ , is similar to that of Spearman, and can be used with any variables that are at least ordinal. It measures the degree to which a relationship always increases or decreases. Let $(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$ be a set of observations of the joint random variables X and Y respectively, such that all the values of x_i and y_i are unique. Any pair of observations (x_i, y_i) and (x_j, y_j) are said to be

concordant if the ranks for both elements agree; that is, if both $x_i > x_j$ and $y_i > y_j$ or if both $x_i < x_j$ and $y_i < y_j$. They are said to be discordant if $x_i > x_j$ and $y_i < y_j$ or if $x_i < x_j$ and $y_i > y_j$. If $x_i = x_j$ or $y_i = y_j$, the pair is neither concordant nor discordant. The Kendall coefficient is defined as:

$$\tau = \frac{D_1 - D_2}{1/2N(N-1)}, \quad (\text{A.3})$$

where D_1 and D_2 are the numbers of concordant and discordant pairs, respectively.

To summarise which coefficient we should use, for interval scale data and if we are interested in linear relationships we use Pearson's coefficient, whilst those of Spearman or Kendall if our interest lies in any decreasing/increasing relationship.

A.2 Mutual information

A pure linear relationship or a linear transformed relationship is required for the application of linear correlations. Thus, these linear correlation coefficients may not be useful in identifying serial dependence if there is nonlinear behaviour in the data [269, 270]. Furthermore, many statistical tests have been used to show that there is evidence of nonlinear dependence and chaotic behaviour in biomedical data. Thus, we also consider a nonlinear measure based on mutual information.

The mutual information of two continuous random variables X and Y can be defined as:

$$I(X, Y) = \int_X \int_Y p(x, y) \log \left(\frac{p(x, y)}{p(x)p(y)} \right) d_y d_x, \quad (\text{A.4})$$

where $p(x, y)$ is the joint probability distribution function of X and Y , and $p(x)$ and $p(y)$ are the marginal probability distribution functions of X and Y , respectively. In the discrete case, we replace the integral by a definite double summation. Intuitively, mutual information measures the information that X and Y share: it measures how much knowing one of these variables reduces our uncertainty about the other.

Mutual information can be expressed as:

$$I(X, Y) = H(X) - H(X \setminus Y) = H(Y) - H(Y \setminus X) = H(X) + H(Y) - H(X, Y), \quad (\text{A.5})$$

where $H(X)$ and $H(Y)$ are the marginal entropies, $H(X \setminus Y)$ and $H(Y \setminus X)$ are the conditional entropies, and $H(X, Y)$ is the join entropy of X and Y . Since $H(X) \geq H(X \setminus Y)$, we have $I(X; Y) \geq 0$; assuming equality, if X and Y are statistically independent. Therefore the mutual information between the vectors of random variables X and Y can be considered as a measure of dependence between these variables, or even better, the statistical correlation of X and Y . The statistics defined in Eq. A.5 satisfy some of the desirable properties of a good measure of dependence [270].

The mutual information takes a value between 0 and infinity, $I(X, Y) \geq 0$, which makes the comparisons difficult between different samples. In this context, [270]–[273], among others, defined and used a standard measure for the mutual information:

$$\xi = (1 - \exp[-2I(X, Y)])^{1/2}. \quad (\text{A.6})$$

This measure can capture the overall dependence, both linear and nonlinear, between X and Y . It varies between 0 and 1 and is thus directly comparable to the linear correlation coefficient based on the relationship between the measures of information theory and variance analysis.

Appendix B

Measures of Accuracy and Statistical Significance of Noise Free Time Series Reconstruction

Root mean square (RMSE) error and mean absolute error (MAE) are common statistical metrics that are used to measure the performance of the methods for the prediction of time series. They are used to measure the differences between predicted values by a model or an estimator and actual observed values. They are also used for image and noise free time series reconstruction. In this thesis, these measures are applied to measure the capability of the proposed approach in the selection of the optimal signal subspace.

B.1 Root mean square error (RMSE)

The root mean square error between the original simulated signal and the reconstructed series after removing the added noise by the proposed approach can be calculated from the following formula:

$$RMSE = \sqrt{\frac{\sum_{t=1}^N (\hat{y}_t - y_t)^2}{N}}. \quad (\text{B.1})$$

Here N denotes the length of the series, \hat{y}_t are the reconstructed values obtained by SSA based on the proposed approach, and y_t are the original signal values before adding noise.

B.2 Mean absolute error (MAE)

We also use mean absolute error (MAE) to examine our results. This measure is used to enhance and confirm the accuracy of the approach. It is also applied to measure the difference between original simulated values (before adding Gaussian error) and the reconstructed noise free series. MAE can be calculated from the following formula:

$$MAE = \frac{1}{N} \sum_{t=1}^N |\hat{y}_t - y_t|, \quad (\text{B.2})$$

where N and \hat{y}_t are as defined above.

Appendix C

Measures for Normality

In statistical analysis, the most used distribution is the normal one [274], which is widely used in medical research and many other fields. There exists a considerable literature for testing normality. In the thesis, we used D’Agostino-Pearson [220] and Shapiro-Wilk [221] as they are the most powerful and used tests.

C.1 The Shapiro-Wilk test for normality

One of the commonly used tests in the theory of testing for normality is the Shapiro and Wilk test (1965). For example, [275] reviewed more than thirty formal procedures that have been proposed for testing normality assumption. In terms of power performance against a broad range of alternatives, the S-W test is the benchmark of omnibus tests for univariate data. Moreover, in 2011, [276] compared the power of four formal tests of normality: the Shapiro-Wilk (S-W) test, the Kolmogorov-Smirnov test, the Lilliefors test and the Anderson-Darling test. The authors concluded that of these considered tests, the S-W test is the most powerful one for all types of distribution and sample sizes. Furthermore, for practical applications, biomedical researchers often prefer to use S-W test as it is informative and easy for them to understand. The test utilizes the null hypothesis principle to test if a

sample x_1, \dots, x_n of n independently and identically distributed (iid) observations come from a normally distributed population.

Let $x_1 \leq \dots \leq x_n$ denote the ordered values of the random sample x_i ($i = 1, \dots, n$),

the S-W statistic can be calculated as follows:

$$\text{S-W} = \frac{(\sum_{i=1}^n a_i x_{(i)})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad (\text{C.1})$$

where $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$ is the sample mean; the constants a_i are obtained by the following formula:

$$A^T = (a_1, \dots, a_n) = \frac{M^T \mathbf{V}^{-1}}{(M^T \mathbf{V}^{-1} \mathbf{V}^{-1} M)^{\frac{1}{2}}}, \quad (\text{C.2})$$

where $M^T = (m_1, \dots, m_n)$, and m_1, \dots, m_n are the expected values of the order statistics of iid random variables generated from the standard normal distribution; and \mathbf{V} is the covariance matrix of those order statistics.

In the S-W test, the null hypothesis is the population normally distributed. Therefore, if the p -value is greater than the chosen alpha level, then the null hypothesis is accepted and we have evidence that the examined data came from a normally distributed population.

C.2 D'Agostino-Pearson normality test

The skewness and kurtosis statistics, and the D'Agostino-Pearson (D-P) [220] statistic that combines these two have been shown to be informative and powerful test for testing normality [277]. There has been a plethora of these tests available, and there is a long history of them [274]. The S-W test, the skewness and kurtosis tests, and the D'Agostino-Pearson test combining these [220] emerge as excellent tests [277]. The S-W and D-P tests share the fine property of being omnibus tests, in that they have good power properties over a broad range of non-normal distributions.

The D'Agostino-Pearson normality test calculates skewness and kurtosis to quantify how far the distribution is from Gaussian in terms of asymmetry and shape. It then computes how these values vary from those expected with Gaussian distribution, and calculates a single p -value from the sum of these discrepancies. The D-P test statistic is:

$$D-P = Z^2(b_1) + Z^2(b_2), \quad (\text{C.3})$$

where $Z^2(b_1)$ and $Z^2(b_2)$ are the standard normal deviates equivalent to observing b_1 (skewness) and b_2 (kurtosis) [278]. The D-P statistic has approximately a chi-squared distribution, with 2 degrees of freedom when the population is normally distributed [277].

References

- [1] Itoh, K. (1996). A noise reduction method for chaotic time series. *Electronics and Communications in Japan*, **79** (12), pp. 1550–1557.
- [2] Kantz, H., and Schreiber, T. (1997). *Nonlinear time series analysis*. Cambridge: Cambridge University Press.
- [3] Abarbanel, H. D. I. (1996). *Analysis of observed chaotic data*. New York: Springer.
- [4] Kapitaniak, T. (1990). *Chaos in systems with noise*. Singapore: World Scientific.
- [5] Hassani, H., Zokaei, M., Rosen, D., Amiri, S., and Ghodsi, M. (2009). Does noise reduction matter for curve fitting in growth curve models?. *Computer Methods and Programs in Biomedicine*, **96**, pp. 173–181.
- [6] Golyandina, N., and Zhigljavsky A. (2013). *Singular spectrum analysis for time series*. Verlag Berlin Heidelberg: Springer Briefs in Statistics. Springer.
- [7] Hassani, H., Heravi, S., Brown, G., and Ayoubkhani, D. (2013). Forecasting before, during, and after recession with singular spectrum analysis. *Journal of Applied Statistics*, **10** (40), pp. 2290–2302.
- [8] Iranmanesh, H., Abdollahzadeh, M., Miranian, A., and Hassani, H. (2013). A developed wavelet-based local neuro fuzzy model for the forecasting of crude oil price. *International Journal of Energy and Statistics*, **1** (3), pp. 171–193.

- [9] Hadad, M., Taibi, H., Hassani, H. (2013). Sea level in the Mediterranean sea: Seasonal adjustment and trend extraction within the framework of SSA. *Earth Science Informatics*, **6** (2), pp. 99–111.
- [10] Chen, Z., Fang, X., Ding, K., Xie, Q., and Xie, L. (2013). Noise reduction of acoustic signals from armored vehicles based on local projective algorithm. *Journal of Information and Computational Science*, **10** (6), pp. 1577–1584.
- [11] Schreiber, T., and Grassberger, P. (1991). A simple noise-reduction method for real data. *Physics Letters A*, **160** (5), pp. 411–418.
- [12] Schreiber, T. (1993). Extremely simple nonlinear noise reduction method. *Physical Review E*, **47** (4), pp. 2401–2404.
- [13] Zheng, H. M., Li, Y. A., and Chen, L. (2011). Noise reduction of ship signals based on the local projective algorithm. *Journal of Northwestern Polytechnical University*, **29** (4), pp. 569–574.
- [14] Lalley, S. P., and Nobel, A. B. Denoising deterministic time series. (2006). *Dynamics*, **3** (4), pp. 259–279.
- [15] Donoho, D. L., and Johnstone, I. M. (1994). Ideal spatial adaption via wavelet shrinkage. *Biometrika*, **81** (3), pp. 425–455.
- [16] Han, M., Liu, Y. H., Xi, J. H., and Guo, W. (2007). Noise smoothing for nonlinear time series using wavelet soft threshold. *IEEE Signal Process Letters*, **14** (1), pp. 62–65.
- [17] Chang, S. G., Yu, B., and Vetterli, M. (2000). Adaptive wavelet thresholding for image denoising and compression. *IEEE Transactions on Image Processing*, **9** (9), pp. 1532–1546.
- [18] Gao, J., Sultan, H., Hu, J., and Tung, W. (2010). Denoising nonlinear time series by adaptive filtering and wavelet shrinkage: A comparison. *IEEE Signal Processing Letters*, **17** (3), pp. 237–240.

- [19] Jako, Z., and Kis, G. (2000). Application of noise reduction to chaotic communications: A case study. *IEEE Transaction on Circuits and Systems-I: Fundamental Theory and Applications*, **47** (12), pp. 1720–1725.
- [20] Tiwari, R. K., and Rajesh, R. (2014). Factorized Hankel optimal singular spectral approach for erratic and noisy seismic signal denoising. *Journal of Applied Geophysics*, **111**, pp. 95–101.
- [21] Soofi, A., and Cao, L. (2002). *Modeling and forecasting financial data: Techniques of nonlinear dynamics*. Boston: Kluwer Academic Publishers.
- [22] Partha, P. K., Sarbani, P., and Goutam, S. (1997). Fetal ECG extraction from single-channel maternal ECG using singular value decomposition. *IEEE Transactions on Biomedical Engineering*, **44**, pp. 51–59.
- [23] Callaerts, D., De Moor, B., Vandewalle, J., Sansen, W., Vantrappen, G., and Janssens, J. (1990). Comparison of SVD methods to extract the foetal electrocardiogram from cutaneous electrode signals. *Medical Biological Engineering and Computing*, **28**, pp. 217–224
- [24] Alonsoa, F. J., Del Castilloa, J. M., and Pintado, P. (2004). Application of singular spectrum analysis to the smoothing of raw kinematic signals. *Journal of Biomechanics*, **38**, pp. 1085–1092.
- [25] Hassani, H., Dionisio, A., and Ghodsi, M. (2010). The effect of noise reduction in measuring the linear and nonlinear dependency of financial markets. *Nonlinear Analysis: Real World Applications*, **11**, PP. 492–502.
- [26] Ghodsi, M., Hassani, H., Sani, S., and Hicks, Y. (2009). The use of noise information for detection of temporomandibular disorder. *Biomedical Signal Processing and Control*, **4**, pp. 79–85.
- [27] Sanei, S., Ghodsi, M., and Hassani H. (2011). An adaptive singular spectrum analysis approach to murmur detection from heart sounds. *Medical Engineering and Physics*, **33** (3), pp. 362–36.

- [28] Broomhead, D., and King, G. (1986). Extracting qualitative dynamics from experimental data. *Physica D*, **20**, pp. 217–236.
- [29] Broomhead, D., and King, G. (1986b). On the qualitative analysis of experimental dynamical systems. In: Sarkar, S., ed. *Nonlinear Phenomena and Chaos*. Bristol: Adam Hilger, pp 113–144.
- [30] Elsner, J. B., and Tsonis, A. A. (1996). *Singular spectrum analysis: A new tool in time series analysis*. New York: Plenum Press.
- [31] Danilov, D., and Zhigljavsky, A. (1997). *Principal components of time series: the Caterpillar method*. St. Petersburg (in Russian): St.Petersburg Press.
- [32] Golyandina, N., Nekrutkin, V., and Zhigljavsky, A. (2001). *Analysis of time series structure: SSA and related techniques*. New York: Chapman and Hall/CRC.
- [33] Sanei, S., and Hassani, H. (2016). *Singular spectrum analysis of biomedical signals*. London: CRC Press.
- [34] Ghodsi, M., Hassani, H., and Sanei, S. (2010). Extracting fetal heart signal from noisy maternal ECG by multivariate singular spectrum analysis. *Statistics and its Interface*, **3** pp. 399–411.
- [35] Tang, TY., Liew, A. WC., and Yan, H. (2010). Periodicity analysis of DNA microarray gene expression time series profiles in mouse segmentation clock data. *Statistics and its Interface*, **3** (3), pp. 413–418.
- [36] Kapl, M., and Mueller, W. G. (2010). Prediction of steel prices: a comparison between a conventional regression model and MSSA. *Statistics and its Interface*, **3**, pp. 369–375.
- [37] Peller, V. (2003). *Hankel operators and their applications*. New York: Springer.
- [38] Hassani, H., and Mahmoudvand, R. (2013). Multivariate singular spectrum analysis: A general view and new vector forecasting approach. *International Journal of Energy and Statistics*, **1**, pp. 55–83.

- [39] Golyandina, N. (2010). On the choice of parameters in singular spectrum analysis and related subspace-based methods. *Statistics and its Interface*, **3** (3), pp. 259–279.
- [40] Khan, M., and Poskitt, D. (2010). Description length based signal detection in singular spectrum analysis. *Monash Econometrics and Business Statistics Working Papers*, 13/10.
- [41] Hassani, H., Mahmoudvand, M., Zokaei, M., and Ghodsi, M. (2012). On the separability between signal and noise in singular spectrum analysis. *Fluctuation and Noise Letters*, **11** (2), pp. 14–25.
- [42] Khan, M., and Poskitt, D. (2013). A note on window length selection in singular spectrum analysis. *Australian and New Zealand Journal of Statistics*, **55** (2), pp. 87–108.
- [43] Sanei, S., Lee, T., and Abolghasemi, V. (2012). A new adaptive line enhancer based on singular spectrum analysis. *IEEE Transactions on Biomedical Engineering*, **59** (2), pp. 428–434.
- [44] Rodriguez-Aragon, L., and Zhigljavsky, A. (2010). Singular spectrum analysis for image processing. *Statistics and its Interface*, **3**, pp. 419–426.
- [45] Celka, P., and Colditz, P. (2002). A computer-aided detection of EEG seizures in infants, a singular-spectrum approach and performance comparison. *IEEE Transactions on Biomedical Engineering*, **49**, pp. 455–462.
- [46] Kouchaki, S., Sanei, S., Arbon, E., and Dijk, D. (2015). Tensor based singular spectrum analysis for automatic scoring of sleep EEG. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, **23** (1), pp. 1–9.
- [47] Sanei, S., and Chambers, J. A. (2007). *EEG signal processing*. New York: Wiley.
- [48] Shmulevich, I., and Dougherty, E. R. (2007). *Genomic signal processing*. Princeton: Princeton University Press.

- [49] Enoka, R. (2008). *Neuromechanics of human movement*. 4th edition. Champaign: Human Kinetics.
- [50] Srinmo, L., and Laguna, P. (2005). *Bioelectrical signal processing in cardiac and neurological applications*. Amsterdam: Elsevier.
- [51] Chang, HH., and Moura, J. *Biomedical signal processing*. in: Kutz, M., ed. *Biomedical Engineering and Design Handbook*. 2nd edition. New York: McGraw Hill, **1**, pp. 559–579.
- [52] Oppelt, A. (2005). *Imaging systems for medical diagnostics*. New York: Wiley-VCH.
- [53] Team Bunraku. (2009). *Perception, decision and action of real and virtual humans in virtual environments and impact on real environments*. France: Rennes. Available from: <http://raweb.inria.fr/rapportsactivite/RA2009/bunraku/bunraku.pdf> [Accessed: 20 December 2015].
- [54] Available from: <https://www.bupaglobal.com> [Accessed: 20 January 2016]
- [55] Lorenz, E. N. (1963). Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, **20**, pp. 130–141.
- [56] Andrzejak, R., Lehnertz, K., Mormann, F., Rieke, C., David, P., and Elger, C. (2001). Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Physical Review E*, **64**, pp. 1–8.
- [57] Harikrishnan, K., Misra, R., Ambika, G., Kembhavi, A. (2006). A non-subjective approach to the GP algorithm for analysing noisy time series. *Physica D*, **215**, pp. 137–145.
- [58] Lehnertz, K., Elger, C., Arnhold, J., and Grassberger, P. (1999). *Chaos in Brain?*. Germany: University of Bonn.
- [59] West, B. J. (2000). *Fractal physiology and chaos in medicine*. Singapore: World Scientific.

- [60] Goldberger, A. L., Bruce, B. J. (1987). Application of non linear dynamics to clinical cardiology. *Annals of the New York Academy of Sciences*, **504**, pp. 195–213.
- [61] Kaplan, D. T., Kohen, R. J. (1990). Is fibrillation chaos?. *Circulation Research*, **67**, pp. 886–900.
- [62] Richter, M., Schreiber, T. and Kaplan, D. T. (1998). Fetal ECG extraction with nonlinear state space projections. *IEEE Transactions on Biomedical Engineering*, **45** (1), pp. 133–137.
- [63] Zhang, X. S., Zhu, Y. S., Wang, Z. Z., and Thakor, N. V. (1999). Detecting ventricular tachycardia and fibrillation by complexity measure. *IEEE Transactions on Biomedical Engineering*, **46**, pp. 548–555.
- [64] Yaylali, I., Koak, H., and Yayakar, P. (1996). Detection of seizures from small samples using nonlinear dynamic system theory. *IEEE Transactions on Biomedical Engineering*, **43**, pp. 743–751.
- [65] Babloyantz, A., and Slazar, J. M. (1985). Evidence of chaos dynamic of brain activity during the sleep cycle. *Physics Letters A*, **111** (3, 2), pp. 152–156.
- [66] Adeli, H., Ghosh-Dastidar, S., and Dadmehr, N. (2007). A wavelet-chaos methodology for analysis of EEGs and EEG sub-bands to detect Seizure and Epilepsy. *IEEE Transactions on Biomedical Engineering*, **54** (2), pp. 205–211.
- [67] Almong, Y., Oz, O., and Akselrod, S. (1999). Correlation dimension estimation: can this non linear description contribute to the characterization of blood pressure control in rats?. *IEEE Transactions on Biomedical Engineering*, **46** (5), pp. 535–547.
- [68] Wagner, C. D., and Persson, P. B. (1995). Non linear chaotic dynamics of arterial blood pressure and renal flow. *American Journal of Physiology*, **268** (37), pp. 621–627.

- [69] Filligoi, G., and Felici, F. (1999). Detection of hidden rhythms in surface EMG signals with a non-linear time-series tool. *Medical Engineering and Physics*, **21**, pp. 439–448.
- [70] Kearney, R. E., and Hunter, J. W. (1998). Non linear identification of stretch reflex dynamics. *Annals of Biomedical Engineering*, **16**, pp. 79–94.
- [71] Hu, X., Wang, Z. Z., and Ren, X. M. (2005). Classification of surface EMG signal with fractal dimension. *Journal of Zhejiang University Science B*, **6** (8), pp. 844–848.
- [72] Dingwell, J., Napolitano, D. F., and Chelidze, D. (2007). A non-linear approach to tracking slow-time scale changes in movement kinematics. *Journal of Biomechanics*, **40**, pp. 1629–1634.
- [73] Song, M., Segala, D. B., Dingwell, J. B., and Chelidze, D. (2009). Slow-time changes in human EMG muscle fatigue states are fully represented in movement kinematics. *Journal of Biomechanical Engineering*, **131** (2), pp. 021004.
- [74] Akay, M., and Mulder, E. J. H. (1998). Effects of maternal alcohol intake on fractal properties in human fetal breathing dynamics. *IEEE Transactions on Biomedical Engineering*, **45** (9), pp. 1097–1103.
- [75] Hoyer, D., Bauer, R., Walter, B., and Zwiener, U. (1998). Estimation of non-linear coupling on the basis of complexity and predictability: a new method applied to cardio-respiratory coordination. *IEEE Transactions on Biomedical Engineering*, **45** (5), pp. 545–552.
- [76] Porrello, A., Soddu, S., Zbilut, J. p., Crescenzi, M., and Giuliani, A. (2004). Discrimination of single amino acid mutations of the p53 protein by means of deterministic singularities of recurrence quantification analysis. *Proteins-Structure Function and Bioinformatics*, **55**, pp. 743–755.
- [77] Grizzi, F., and Chiriva Internati, M. (2005). The complexity of anatomical systems. *Theoretical Biology and Medical Modelling*, **2**. Available from: <http://www.tbiomed.com/content/2/1/26> [Accessed 23 December 2015].

-
- [78] Bianciardi, M., Sirabella, P., Hagberg, G. E., Giuliani, A., Zbilut, J. P., and Colosimo, A. (2007). Model-free analysis of brain fMRI data by recurrence quantification. *NeuroImage*, **37** (2), pp. 489–503.
- [79] Kostelich, E. J., and Yorke, J. A. (1988). Noise reduction in dynamical systems. *Physical Review. A*, **38** (3), pp. 1649–1652.
- [80] Farmer, J. D., and Sidorowich, J. J. (1991). Optimal shadowing and noise reduction. *Physica D*, **47**, pp. 373–392.
- [81] Cawley R., and Hsu, G. H. (1992). SNR performance of a noise reduction algorithm applied to coarsely sampled chaotic data. *Physics Letters A*, **166** (3, 4), pp. 188–196.
- [82] Hammel, S. M. A noise reduction method for chaotic systems. (1990). *Physics Letters A*, **148** (8, 9), pp. 421–428.
- [83] Sauer, T. (1992). A noise reduction method for signals from nonlinear systems. *Physica D: Nonlinear Phenomena*, **58**, pp. 193–201.
- [84] Vautard, R., Yiou, P., and Ghil, M. (1998). Singular spectrum analysis: a toolkit for short noisy chaotic signals. *Physica D: Nonlinear Phenomena*, **58**, pp. 95–126.
- [85] Davies, M. (1994). Noise reduction schemes for chaotic time series. *Physica D: Nonlinear Phenomena*, **79**, pp. 174–192.
- [86] Takens, F. (1981). Detecting strange attractors in turbulence. *in*: Rand, D., and Young, L.S., eds. *Dynamical Systems and Turbulence*. Berlin: Springer, **898**, pp. 366–381.
- [87] Kostelich, E. G., and Schreiber, T. (1993). Noise reduction in chaotic timeseries data: a survey of common methods. *Physical Review E*, **48**, pp. 1752–1763.
- [88] Broomhead, D. S., Huke, J. P., and Jones, R. (1995). Signals in chaos: a method for the cancellation of deterministic noise from discrete signals. *Physica D: Nonlinear Phenomena*, **38**, pp. 423–432.

- [89] Provenzale, A., Smith, L., Vio, R., and Murante, G. (1992). Distinguishing between low-dimensional dynamics and randomness in measured time series. *Physica D: Nonlinear Phenomena*, **58**, pp. 31–49.
- [90] Dennis, B., Desharnais, R. A., and Cushing, J. M. (2003). Can noise induce chaos?. *Oikos*, **102**, pp. 329–340.
- [91] Morris, W. F. (1990). Problems in detecting chaotic behavior in natural populations by fitting simple discrete models. *Ecology*, **71**, pp. 1849–1862.
- [92] Ellner, S., and Turchin, P. (1995). Chaos in a noisy world: New methods and evidence from time series analysis. *The American Naturalist*, **145**, pp. 343–375.
- [93] Turchin, P. (1995). Chaos in microtine populations. *Proceedings of the Royal Society. B*, **262**, pp. 357–361.
- [94] Turchin, P. (2003). *Complex population dynamics: A theoretical/empirical synthesis*. Princeton: Princeton University Press.
- [95] Ellner, S., and Turchin, P. (2005). When can noise induce chaos and why does it matter: a critique. *Oikos*, **111**, pp. 620–631.
- [96] Scheuring, I., and Domokos, G. (2007). Only noise can induce chaos in discrete populations. *Oikos*, **116**, pp. 361–366.
- [97] Hassani, H. (2010). Singular spectrum analysis based on the minimum variance estimator. *Nonlinear Analysis: Real World Applications*, **11** (3), pp. 2065–2077.
- [98] Hassani, H., and Zhigljavsky, A. (2009). Singular Spectrum Analysis: Methodology and Application to Economics Data. *Journal of System Science and Complexity*, **22**, pp. 372394.
- [99] Vautard M, Ghil M(1989) Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series. *Physica D* 35:395-424
- [100] Allen, M., and Smith, L. (1996). Monte Carlo SSA: Detecting irregular oscillations in the presence of colored noise. *J Clim*, **9** (12), pp. 3373–3404.

- [101] Fraedrich, K. (1986). Estimating dimensions of weather and climate attractors. *J Atmos Sci*, **43**, pp. 419-432.
- [102] Nekrutkin, V. (2010). Perturbation expansions of signal subspaces for long signals. *Stat Interface*, **3**, pp. 297-319.
- [103] Usevich, K. (2010). On signal and extraneous roots in singular spectrum analysis. *Stat Interface*, **3** (3), pp. 281-295.
- [104] Moskvina, V., and Zhigljavsky, A. (2003). An algorithm based on singular spectrum analysis for change-point detection. *Commun Stat Simul Comput*, **32**, pp. 319-352.
- [105] . Golyandina, N., Pepelyshev, A., and Steland, A. (2012). New approaches to nonparametric density estimation and selection of smoothing parameters. *Comput Stat Data Anal* **56** (7), pp. 2206-2218.
- [106] Granger, CWJ. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, **37** (3), pp. 424-438.
- [107] Hassani, H., Heravi, H., and Zhigljavsky, A. (2009). Forecasting European industrial production with singular spectrum analysis. *International Journal of Forecasting*, **25** (1), pp. 103–118.
- [108] Golyandina, N., and Usevich, K. (2010). 2D-Extension of singular spectrum analysis: Algorithm and elements of theory. *In: Olshevsky, V., and Tyrtshnikov, E., eds. Matrix methods: Theory, algorithms and applications*, World Scientific Publishing, pp. 449-473.
- [109] Alonsoa, F. J., Del Castilloa, J. M., and Pintado, P. (2005). Application of singular spectrum analysis to the smoothing of raw kinematic signals. *Journal of Biomechanics*, **38**, pp. 1085–1092.
- [110] Zhigljavsky, A. (2010). Special issue on theory and practice in singular spectrum analysis of time series. *Statistics and its Interface*, **3** (3), pp. 255—258.

- [111] Azulay, DO., Brain, P., and Sultana, SR. (2011). Characterisation of very low frequency oscillations in laser Doppler perfusion signals with a singular spectrum analysis. *Microvasc Res*, **81** (3), pp. 239-244.
- [112] Hassani, H. (2007). Singular spectrum analysis: Methodology and comparison. *Journal of Data Science*, **5**, pp. 239–257.
- [113] Patterson, K., Hassani, H., Heravi, S., and Zhigljavsky, A. (2011). Multivariate singular spectrum analysis for forecasting revisions to real-time data. *J of Appl Stat*, **38** (10), pp. 2183-2211.
- [114] Weare, BC., and Nasstrom, JS. (1982). Examples of extended empirical orthogonal function analyses. *Mon Weather Rev*, **110** (6), pp. 481–485.
- [115] Colebrook, JM. (1978). Continuous plankton records-zooplankton and environment, northeast Atlantic and North Sea. *Oceanol Acta*, **1**, pp. 9-23.
- [116] Kondrashov, D., and Ghil, M. (2006). Spatio-temporal filling of missing points in geophysical data sets. *Nonlinear Process, Geophys*, **13** (2), pp. 151-159.
- [117] Alexandrov, T., Golyandina, N., and Spirov, A. (2008). Singular spectrum analysis of gene expression profiles of early Drosophila embryo: exponential-in-distance patterns. *Research Letters in Signal Processing*, **2008**, pp. 1–5.
- [118] Hassani, H., and Ghodsi, Z. (2014). Pattern recognition of gene expression with singular spectrum analysis. *Medical Sciences*, **2** (3), pp. 127-139.
- [119] Mahecha, MD., Frst, LM., Gobron, N., Lange, H. (2010). Identifying multiple spatiotemporal patterns: A refined view on terrestrial photosynthetic activity. *Pattern Recogn Lett*, **31** (14), pp. 2309-2317.
- [120] De Carvalho, M., Rodrigues, PC., and Rua, A. (2012). Tracking the US business cycle with a singular spectrum analysis. *Econ Lett*, **114** (1), pp. 32-35.
- [121] Parra, L. C., and Sajda, P. (2003). Blind source separation via generalized eigenvalue decomposition. *Journal of Machine Learning Research* **4**, pp. 1261–1269.

- [122] Calhoun, V., Liu, J., and Adali, T. (2009). A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data. *NeuroImage*, **45**, pp. 163–172.
- [123] De Vos, M., Vergult, A., De Lathauwer, L., De Clercq, W., Van Huffel, S., Dupont, P., Palmieri, A., and Van Paesschen, W. (2007). Canonical decomposition of ictal scalp EEG reliably detects the seizure onset zone. *NeuroImage*, **37**, pp. 844–854.
- [124] Li, Y., Cichocki, A., and Amari, S. (2003). Sparse component analysis for blind source separation with less sensors than sources. In: *proceeding, the 4th international symposium on independent component analysis and blind signal separation(ICA)*, Kyoto, Japan, April 2003. Nara: Riken Brain Science Institute, pp. 89–94. Available from: <http://www.bsp.brain.riken.jp/publications/2003/ICA03LiAmariCich.pdf> [Accessed 28 December 2015].
- [125] Zhou, P., and Kuiken, T. (2006). Eliminating cardiac contamination from myoelectric control signals developed by targeted muscle reinnervation. *Physiological Measurement*, **27**, pp. 1311–1327.
- [126] Lewicki, M. (1998). A review of methods for spike sorting: The detection and classification of neural action potentials. *Computation and Neural Systems*, **9**, pp. 53–78.
- [127] Niedermeyer, E. (1997). Alpha rhythms as physiological and abnormal phenomena. *International Journal of Psychophysiology*, **26**, pp. 31–49.
- [128] Mistlberger, R. E., Bergmann, B. M., and Rechtschaffen, A. (1987). Relationships among wake episode lengths, contiguous sleep episode lengths, and electroencephalographic delta waves in rats with suprachiasmatic nuclei lesions. *Sleep*, **10**, pp. 12–24.
- [129] Pereira, W. C. A., Bridal, S. L., Coron, A., and Laugier, P. (2004). Singular spectrum analysis applied to backscattered ultrasound signals from in vitro

- human cancellous bone specimens. *IEEE Transactions on Ultrasonics Ferro-electrics and Frequency Control*, **51** (3), pp. 302–312.
- [130] Teixeira, A. R., Tome, A. M., Lang, E. W., Gruber, P., and Martins da Silva, A. (2005). On the use of clustering and local singular spectrum analysis to remove ocular artifacts from electroencephalograms. *In: Proceeding of the International Joint Conference on Neural Networks*, Montreal, Canada, 31 July-4 August 2005. Montreal: IEEE. **4**, pp. 2514–2519. Available from: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1556298> [Accessed 5 January 2016].
- [131] Teixeira, A. R., Tome, A. M., Bohm, M., Puntonet, C. G., and Lang, E. W. (2009). How to apply nonlinear subspace techniques to univariate biomedical time series. *IEEE Transactions on Instrumentation and Measurement*, **58** (8), pp. 2433–2443.
- [132] Sanei, S., Ghodsi, M., and Hassani, H. (2011). A constrained singular spectrum analysis approach to murmur detection from heart sounds. *Medical Engineering and Physics*, **33** (3), pp. 362–367.
- [133] Sanei, S., and Hosseini-Yazdi, A. R. (2011). Extraction of ECG from single channel EMG signal using constrained singular spectrum analysis. *In: 17th International Conference on Digital Signal Processing*, Corfu, Greece, 6-8 July 2011. Corfu: IEEE. pp. 1–4. Available from: <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6004876> [Accessed 5 January 2016].
- [134] Ghaderi, F., Mohseni, H. R., and Sanei, S. (2011). Localizing heart sounds in respiratory signals using singular spectrum analysis. *IEEE Transactions on Biomedical Engineering*, **58** (12), pp. 3360–3367.
- [135] McNeill, C. (1993). *Temporomandibular disorders: Guidelines for classification, assessment and management*. Chicago: Quintessence.
- [136] Okeson, J. P. (1996). Temporomandibular disorders in the medical practice. *The Journal of Family Practice*, **43** (4), pp. 347–356.

- [137] Camps, G., Martinez, M. and Sofia, E. (2001). Fetal ECG extraction using an FIR neural network. *IEEE, Computers in Cardiology*, **28**, pp. 249–252.
- [138] Abboud, S., Alaluf, A., Einav, S. and Sadeh, D. (1992). Realtime abdominal fetal ECG recording using a hardware correlator. *Computers in Biology and Medicine*, **22**, pp. 325–335
- [139] Kam, A. and Cohen, A. (1999). Detection of fetal ECG with IIR adaptive filtering and genetic algorithms. In: *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Phoenix, AZ, 15-19 March 1999. New York: IEEE. pp. 2335-2338. Available from: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=758406 [Accessed 25 January 2016].
- [140] Mooney, D. M., Groome, L. J., Bentz, L. S. and Wilson, J. D. (1995). Computer algorithm for adaptive extraction of fetal cardiac electrical signal. In: *proceedings of the ACM symposium on applied computing*, New York 1995. New York: ACM. pp. 113–117. Available from: <http://dl.acm.org/citation.cfm?id=315932> [Accessed 25 January 2016].
- [141] Martinez, M., Sofia, E., Calpe, J., Guerrero, J. F. and Magdalena, J. R. (1997). Application of the adaptive impulse correlated filter for recovering fetal electrocardiogram. *Computers in Cardiology*, **24**, pp. 9–12.
- [142] Longini, R., Reichert, T., Cho, J. and Crowley, J. (1997). Near orthogonal basis functions: A real time fetal ECG technique. *IEEE Transactions on Biomedical and Engineering*, **24**, pp. 29–43.
- [143] Azad, K. A. K. (2000). Fetal QRS complex detection from abdominal ECG: A fuzzy approach. In: *proceeding IEEE Nordic signal processing symposium*, Sweden 2000. Sweden: IEEE. pp. 275–278. Available from: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.463.9948> [Accessed 27 January 2016].

- [144] De Lathauwer, L., De Moor, B. and Vandewalle, J. (2000). Fetal electrocardiogram extraction by blind source subspace separation. *IEEE Transactions on Biomedical Engineering*, **47** (5), pp. 567-572.
- [145] Zarzoso, V., and Nandi, A. K. (2001). Noninvasive fetal electrocardiogram extraction: blind separation versus adaptive noise cancellation. *IEEE Transactions on Biomedical Engineering*, **48** (1), pp. 12-18.
- [146] Barros, A. K. (2002). Extracting the fetal heart rate variability using a frequency tracking algorithm. *Neurocomputing*, **49**, pp. 279-288.
- [147] Ibahimy, M. I., Ahmed, F., Mohd Ali, M. A. and Zahedi, E. (2003). Real time signal processing for fetal heart rate monitoring. *IEEE Transactions on Biomedical Engineering*, **50**, pp. 258-262
- [148] Khamene, A. and Negahdaripour, S. (2000). A new method for the extraction of fetal ECG from the composite abdominal signal. *IEEE Transaction on Biomedical Engineering*, **47**, pp. 507-516.
- [149] Mochimaru, F., Fujimoto, F. and Ishikawa, Y. (2002). Detecting the fetal electrocardiogram by wavelet theory-based methods. *Progress in Biomedical Research*, **7**, pp. 185-193.
- [150] Karvounis, E. C., Papaloukas, C., Fotiadis, D. I. and Michalis, L. K. (2004). Fetal heart rate extraction from composite maternal ECG using complex continuous wavelet transform. In: *proceeding, computers in cardiology*, Chicago 19-22 September 2004. Chicago: IEEE.
- [151] Kotas, M. (2007). Projective filtering of time-aligned beats for foetal ECG extraction. *Bulletin of the Polish Academy of Science, Technical Science*, **55**, pp. 331-339.
- [152] Sebastian, S., and Rathnakara, S. (2013). Separation of heart sound artifact from respiratory signals using singular spectrum based advanced line enhancer. *International Journal on Advanced Computer Theory and Engineering*, **2** (4), pp. 107-112.

-
- [153] Thuraisingham, R. A. (2013). Use of SSA and MCSSA in the analysis of cardiac RR time series. *Journal of Computational Medicine*, **2013**, pp. 1–8.
- [154] Itoh, N., and Kurths, J. (2011). Forecasting of EEG data by singular spectrum analysis. *Pacific Science Review*, **13** (3), pp. 131–139.
- [155] Allison, D. B., Cui, X., Page, G. P., and Sabripour M. (2006). Microarray data analysis: from disarray to consolidation and consensus. *Nature Reviews Genetics*, **7** (1), pp. 55–65.
- [156] Goeman, J. J., and Bhlmann, P. (2007) Analyzing gene expression data in terms of gene sets: methodological issues. *Bioinformatics*, **23** (8), pp. 980–987.
- [157] VanGuilder, H. D., Vrana, K. E., and Freeman, WM. (2008). Twenty-five years of quantitative PCR for gene expression analysis. *Biotechniques*, **44** (5), pp. 619–626.
- [158] Bustin, S. A. (2014). The reproducibility of biomedical research: sleepers awake!. *Biomolecular Detection and Quantification*, **2**, pp. 35–42.
- [159] Huggett, J., Grady, J., and Bustin, S. (2014). How to make mathematics biology’s next and better microscope. *Biomolecular Detection and Quantification*, **1** (1), pp. 1–3.
- [160] Kruglyak, L., Daly, M. J., Reeve, M. P., and Lander E. S. Parametric and nonparametric linkage analysis: a unified multipoint approach. *The American Journal of Human Genetics*, **58** (6), pp. 1347–1363.
- [161] Grimm, O., Coppey, M., Wieschaus, E. (2010). Modelling the bicoid gradient. *Development*, **137** (14), pp. 2253–2264.
- [162] Du, L., Wu, S., Liew, A. WC, Smith, D., Yan, H. (2008). Spectral analysis of microarray gene expression time series data of *Plasmodium falciparum*. *International Journal of Bioinformatics Research and Applications*, **4** (3), pp. 337–349.

- [163] Tang, TY., and Yan, H. (2012). Noise reduction in microarray gene expression data based on spectral analysis. *International Journal of Machine Learning and Cybernetics*, **3** (1), pp. 51-57.
- [164] Holloway, D. M., Harrison, L. G, Kosman, D., Vanario, A. C. E., and Spirov A. V. (2006). Analysis of pattern precision shows that *Drosophila* segmentation develops substantial independence from gradients of maternal gene products. *Developmental Dynamics*, **235** (11), pp. 2949-2960.
- [165] Gregor, T., Wieschaus, E. F., McGregor, A. P., Bialek, W., and Tank, D. W. Stability and nuclear dynamics of the bicoid morphogen gradient. *Cell*, **130** (1), pp. 141-152.
- [166] Hilfinger, A., and Paulsson, J. (2011). Separating intrinsic from extrinsic fluctuations in dynamic biological systems. *Proceedings of the National Academy of Sciences*, **108** (29), pp. 12167-12172.
- [167] Klebanov, L., and Yakovlev, A. (2007). How high is the level of technical noise in microarray data?. *Biology Direct*, **2**, pp. 1–9.
- [168] Hassani, H., and Ghodsi, Z. (2015). A glance at the applications of singular spectrum analysis in gene expression data. *Biomolecular Detection and Quantification*, **4**, pp. 17–21.
- [169] Surkova, S., Kosman, D., Koslov, K., Manu, E., Myasnikova, A., Samsonova, A., and Reinitz, J. (2008). Characterization of the *Drosophila* segment determination morphome. *Developmental Biology*, **313** (2), pp. 844-862.
- [170] Lopes, F. J. P., Vieira, F. M. C., Holloway, D. M., Bisch, P. M., Spirov, A. V. (2008). Spatial bistability generates hunchback expression sharpness in the *Drosophila* embryo. *PLoS Computational Biology*, **4** (9): e1000184. doi: 10.1371/journal.pcbi.1000184.
- [171] Liew A. WC., Law, NF., Cao, XQ., and Yan, H. (2009). Statistical power of Fisher test for the detection of short periodic gene expression profiles. *Pattern Recognition*, **42** (4), pp. 549-556.

- [172] Liew, A. WC., Xian, J., Wu, S., Smith D., and Yan H. (2007). Spectral estimation in unevenly sampled space of periodically expressed microarray time series data. *BMC Bioinformatics*, **8** (137), PP. 1–19..
- [173] Yeung, L. K., Szeto, L. K., Liew, A. WC., and Yan, H. (2004). Dominant spectral component analysis for transcriptional regulations using microarray time-series data. *Bioinformatics*, **20** (5), pp. 742-749.
- [174] Bozdech, Z., Llinas, M., Pulliam, B. L., Wong, E. D., Zhu, J. C., and DeRisi, J. L. (2003). The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *PLoS Biology*, **1** (1), pp. 1-16.
- [175] Tang, TY., and Yan, H. (2010). Identifying periodicity of microarray gene expression profiles by autoregressive modeling and spectral estimation. *In: Ninth International Conference on Machine Learning and Cybernetics*, Qingdao 11-14 July 2010. China, Qingdao: IEEE, **306**, pp. 3062–3066.
- [176] Tang, TY., Liew, A. WC., and Yan, H. (2010). Analysis of mouse periodic gene expression data based on singular value decomposition and autoregressive modeling. *In: Proceeding of the International MultiConference of Engineers and Computer Scientists*, Hong Kong 17-19 March 2010. Hong Kong: International Association of Engineers.
- [177] Liew, A. WC., and Yan, H. (2009). Reliable detection of short periodic gene expression time series profiles in DNA microarray data. *In: International Conference SMC*, San Antonio, TX, USA 11-14 October 2009. San Antonio: IEEE, pp. 4274–4279. a set theoretic framework and biological consideration. *Nucleic Acids Research*, **34** (5), pp. 1608-1619
- [178] Vikalo, H., Hassibi, B., and Hassibi, A. (2008). Modeling and estimation for real-time microarrays. *IEEE journal of selected topics in signal processing*, **2** (3), pp. 286-296.
- [179] Rau, A., Jaffrezic, F., Foulley, JL., and Doerge, R. W. (2010). An empirical Bayesian method for estimating biological networks from temporal microarray data. *Statistical Applications in Genetics and Molecular Biology*, **9**, pp. 1-9.

- [180] Bremer, M., and Doerge, R. W. (2009). The KM-algorithm identifies regulated genes in time series expression data. *Advances in Bioinformatics*, **2009**, pp. 1–10.
- [181] Golyandina, N., Holloway, D. M., and Lopesc, F. J. P. (2012). Measuring gene expression noise in early *Drosophila* embryos: nucleus-to-nucleus variability. *Procedia Computer Science*, **9**, pp. 373–382.
- [182] Brocwell, P., and Davis, R. (1986). *Time series: Theory and methods*. New York: Springer.
- [183] Chow, P., Khasminskii, R., and Liptser, R. (1997). Tracking of signal and its derivatives in Gaussian white noise. *Stochastic Processes and their Applications*, **69**, pp. 259–273.
- [184] Whitney, H., and Takens, F. (2006). *Embedding theorems*. Available from: http://scholarpedia.org/article/Attractor_reconstruction [Accessed: 25 April 2014].
- [185] Sauer, T., Yorke, J. A., and Casdagli, M. (1991). Embedology. *Journal of Statistical Physics*, **65** (3,4), pp. 579–616.
- [186] Hassani, H., Soofi, A., and Zhigljavsky, A. (2013). Predicting inflation dynamics with singular spectrum analysis. *Journal of the Royal Statistical Society-series A*, **176** (3), pp. 743–760.
- [187] Hassani, H., and Thomakos, D. (2010). A Review on singular spectrum analysis for economic and financial time series. *Statistics and its Interface*, **3** (3), pp. 377–397.
- [188] Chugunov, V. N. (2011). On the parametrization of classes of normal Hankel matrices. *Computational Mathematics and Mathematical Physics*, **51** (11), pp. 1823–1836.
- [189] Pastur, L.A. (2005). A simple approach to the global regime of Gaussian ensembles of random matrices. *Ukrainian Mathematical Journal*, **57** (6), pp. 936–966.

- [190] Wigner, E. P. (1955). Characteristic vectors of border matrices with infinite dimensions. *Annals of Mathematics*, **62**, pp. 548–564.
- [191] Bai, Z., Fang, Z., Liang, YC. (2014). *Spectral theory of large dimensional random matrices and its applications to wireless communications and finance statistics: random matrix theory and its applications*. Singapore: World Scientific Publishing.
- [192] Mehta, M. L. (1991). *Random matrices*. New York: Academic Press.
- [193] Mehta, M. L. (2004). *Random Matrices*. Amsterdam: Elsevier/Academic Press.
- [194] Bai, Z. D., and Silverstein, J. W. (2006). *Spectral analysis of large dimensional random matrices*. Beijing: Science Press.
- [195] Wigner, E. P. (1958). On the distribution of the roots of certain symmetric matrices. *Annals of Mathematics*, **67**, pp. 325–327.
- [196] Arnold, L. (1967). On the asymptotic distribution of the eigenvalues of random matrices. *Journal of Mathematical Analysis and Applications*, **20**, pp. 262–268.
- [197] Arnold, L. (1971). On Wigner’s semicircle law for the eigenvalues of random matrices. *Z. wahrsch. verw. gebiete*, **19**, pp. 191–198.
- [198] Bai, Z. D., and Yin, Y. Q. (1988a). A convergence to the semicircle law. *Annals of Probability*, **16** (2), pp. 863–875.
- [199] Marchenko, V. A., Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Math sbornik*, **1** (4), pp. 507–536.
- [200] Pastur, L. A. (1972). On the spectrum of random matrices. *Theoretical and mathematical physics*, **10**, pp. 67–74.
- [201] Pastur, L. A. (1973). Spectra of random self-adjoint operators. *Russian Mathematical Surveys*, **28** (1).

- [202] Bai, Z. D., Yin, Y. Q., and Krishnaiah, P. R (1986). On limiting spectral distribution of product of two random matrices when the underling distribution is isotropic. *Journal of Multivariate Analysis*, **19**, pp. 189–200.
- [203] Grenander, U., and Silverstein, J. W. (1977). Spectral analysis of networks with random topologies. *Society for Industrial and Applied Mathematics*, **32** (2), pp. 499–519.
- [204] Jonsson, D. (1982). Some limit theorems for the eigenvalues of a sample covariance matrix. *Journal of Multivariate Analysis*, pp. 1–38.
- [205] Wachter, K. W. (1978). The strong limits of random matrix spectra for sample matrices of independent elements. *Annals of Probability*, **6** (1), pp. 1–18.
- [206] Yin, Y. Q. (1986). Limiting spectral distribution for a class of random matrices. *Journal of Multivariate Analysis*, **20** (1), pp. 50–68.
- [207] Bai, Z. D., Yin, Y. Q., and Krishnaiah, P. R (1987). On the limiting empirical distribution function of the eigenvalues of a multivariate F matrix. *Theory of Probability and its Applications*, **32** (3), pp. 490–500.
- [208] Wachter, K. W. (1980). The limiting empirical measure multiple discriminant ratios. *Annals of Statistics* , **8**, pp. 937–957.
- [209] Yin, Y. Q., and Krishnaiah, P. R. (1983). A limit theorem for the eigenvalues of products of two random matrices. *Journal of Multivariate Analysis*, **13**, pp. 489–507.
- [210] Edelman, A., and Rao, N. R. (2005). Random matrix theory. *Acta Numerica*, **14**, pp. 233–297.
- [211] Anderson , G. W., Guionnet, A., and Zeitouni, O. (2010). *An introduction to random matrices*. Cambridge: Cambridge University Press.
- [212] Wishart, J. (1928). Generalized product moment distribution in samples. *Biometrika*, **20** (12), pp. 32–52.

- [213] Mardia, K., Kenet. J, and Dibby. J. (1995). *Multivariate analysis*. London: Academic Press.
- [214] Anderson, T. W. (1963). Asympototic theory for principal component analysis. *Annals of Mathematical Statistics*, **34**, pp. 122–148.
- [215] Pastur, L. A., and Shcherbina, M. (2011). *Eigenvalue distribution of large random matrices*. American Mathematical Society.
- [216] Friedberg, S., Insel, A., and Spence, L. (2002). *Linear algebra*. New York: Pearson.
- [217] Mahmoudvand, R., and Zokaeo, M. (2012). On the singular values of the Hankel matrix with application in singular spectrum analysis. *Chilean Journal of Statistics*, **3** (1), pp. 43-56.
- [218] Bhatia, R. (1997). *Matrix analysis*. New York: Springer.
- [219] Bulmer, M. (1979). *Principles of statistics*. New York: Dover.
- [220] D’Agostino, R. B., and Pearson, E. S. (1973). Testing for departures from normality. 1. fuller empirical results for the distribution of b_2 and b_1 . *Biometrika*, **60**, pp. 613–622.
- [221] Shapiro, S. S., and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, **52** (3,4), pp. 591–611.
- [222] Booeing, G. (2016). Visual analysis of nonlinear dynamical systems: chaos, fractals, self-similarity and the limits of prediction. *Systems*, **4** (4), pp. 37.
- [223] Brock, W. (1986). Distinguishing random and deterministic systems: abridged version. *Journal of Economic Theory*, **40**, pp. 168–195.
- [224] Poincare, Jules Henri. (1890). Sur le problme des trois corps et les quations de la dynamique. Divergence des sries de M. Lindstedt. *Acta Mathematica*, **13**, pp. 1-27.
- [225] Smale, S. (1960). Morse inequalities for a dynamical system. *Bulletin of the American Mathematical Society*, **66**, pp. 43-49.

- [226] Mandelbrot, B. (1967). How long is the coast of Britain? Statistical self-Similarity and fractional dimension. *Science*, bf 156, pp. 636-638.
- [227] Mandelbrot, B. (1982). *The fractal geometry of nature*. New York: Macmillan.
- [228] Gleick, J. (1987). *Chaos: making a new science*. New York: Viking Press.
- [229] Bozoki, Z. (1997). Chaos theory and power spectrum analysis in computerized cardiotocography. *European Journal of Obstetrics and Gynecology and Reproductive Biology*, **71**, (2), pp. 163-168.
- [230] May, R. M. (1974). Biological populations with nonoverlapping generations—stable points, stable cycles and chaos. *Science*, **186**, pp. 645–647.
- [231] May, R. M. (1976). Simple mathematical models with very complicated dynamics. *Nature*, **261**, pp. 459–467.
- [232] Kumar, A., and Hegde, B.M. (2012). Chaos theory: Impact on and applications in medicine. *Nitte University Journal of Health Science*, **2** (4), pp. 2249–2271.
- [233] Casseleggio, A., and Carana, A. (1995). Correlation dimension estimation from electrocardiograms. *Chaos, Solitons and Fractals*, **5**, pp. 713-726.
- [234] Kumar, J., and Kumar, A. (2002). Chaos science and weight bearing foci in the biped. In: *National Conference of the Anatomical Society of India*. Available from: <http://medind.nic.in/jae/t05/i1/jaet05i1p32.pdf> [Accessed: 25 December 2015].
- [235] Poon, C-S., and Merrill, C. (1997). Decrease of cardiac chaos in congestive heart failure. *Nature*, **389**, pp. 492–495.
- [236] Zhao, Y., Sun, J., and Small, M. (2008). Evidence consistent with deterministic chaos in human cardiac data: surrogate and nonlinear dynamical modeling. *International Journal of Bifurcation and Chaos*, **18** (1), pp. 141–160.
- [237] Babloyantz, A., and Destexhe, A. (1988). Is the normal heart a periodic oscillator?. *Biological Cybernetics*, **58**, pp. 203–211.

- [238] Ravelli, F., and Antolini, R. (1992). Complex dynamics underlying the human electrocardiogram. *Biological Cybernetics*, **67**, pp. 57-65.
- [239] Govindan, R. B., Narayanan, K., and Gopinathan, M. S. (1998). On the evidence of deterministic chaos in ECG: Surrogate and predictability analysis. *Chaos*, **8**, pp. 495-502.
- [240] Sugihara, G., and May, R.M. (1990). Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature*, **344**, pp. 734-741.
- [241] Bttcher, F., Peinke, J., Kleinhans, D., Friedrich, R., Lind, P. G., and Haase, M. (2006). Reconstruction of complex dynamical systems affected by strong measurement noise. *Physical Review Letters*, **97**, pp. 603-607.
- [242] Friedrich, R., Siegert, S., Peinke, J., Lck, S., Siefert, M., Lindemann, M., Raethjen, J., Deuschl, G., and Pfister, G. (2000). Extracting model equations from experimental data. *Physics Letters A*, **271** (3), pp. 217- 222.
- [243] Gao, J. B., Hwang, S. K., and Liu, J. M. (1999). When can noise induce chaos? *Physical Review Letters*, **82**, pp. 1132-1135.
- [244] Aulbach, B., and Kieninger, B. (2001). On three definitions of chaos. *Nonlinear Dynamics and Systems Theory*, **1**, pp. 23-37.
- [245] Li, T.J., and Yorke, J.A. (1975). Period three implies chaos. *The American Mathematical Monthly*, **82**, pp. 985-992
- [246] Henon, M. (1976). A two-dimensional mapping with a strange attractor. *Communications in Mathematical Physics*, **50**, pp. 68-77.
- [247] Smirnov, K. (1933). Sulla determinazione empirica di una legge di distribuzione. *Giornale Dell'Istituto Italiano Degli Attuari*, **4**, pp. 83-91.
- [248] <http://www.cs.ucr.edu/~eamonn/discords/> [Accessed 15 November 2014].

- [249] Keogh, E., Lin, J., and Fu, AW. (2005). Hot sax: efficiently finding the most unusual time series subsequence. *In: Proceedings of the 5th IEEE international conference on data mining*, Houston, TX, 27-30 November 2005, pp. 226-233.
- [250] Welvaert, M., and Rosseel, Y. (2013). On the definition of signal-to-noise ratio and contrast-tonoise ratio for fMRI data. *PLoS One*, **8**, (11), pp. 1–10.
- [251] Kannathal, N., Choo, M. L., and Sadasivana, P. K. (2005). Entropies for detection of epilepsy in EEG. *Computer Methods and Programs in Biomedicine*, **80** (3), pp. 187–194.
- [252] Porcher, A., and Dostatni, N. (2010). The bicoid morphogen system. *Current Biology*, **20** (5), PP. 249–254.
- [253] Houchmandzadeh, B., Wieschaus, E., and Leibler, S. (2002). Establishment of developmental precision and proportions in the early Drosophila embryo. *Nature*, **415**, pp. 798–802.
- [254] <http://urchin.spbcas.ru/flyex/> [Accessed: 5 February 2014]
- [255] Lahiri, K., Vaughan, D. R., and Wixon, B. (1995). Modeling SSA’s sequential disability determination process using matched SIPP data. *Social Security Bulletin*, **58** (4), pp. 3–42.
- [256] Guerrero-Mosquera, C., Trigueros, A. M., and Navia-Vazquez, A. (2012). EEG signal processing for epilepsy. *EpilepsyHistological, Electroencephalographic and Psychological Aspects*, **3**, pp. 49–74.
- [257] Khatwani, P., and Tiwari, A. (2013). A survey on different noise removal techniques of EEG signals. *International Journal of Advanced Research in Computer and Communication Engineering*, **2** (2). pp. 1091–1095.
- [258] Kang, D., and Zhizeng, L. (2012). A method of denoising multi-channel EEG signals fast based on PCA and DEBSS algorithm. *In: International Conference on Computer Science and Electronics Engineering*, 23-25 March 2012, **3**, pp. 322-326.

- [259] Babu, P. A. and Prasad, K. V. (2011). Removal of ocular artifacts from EEG signals using adaptive threshold PCA and wavelet transforms. *In: International Conference on Communication Systems and Network Technologies, IEEE Computer Society*, pp. 572–575.
- [260] Karhunen, J., and Joutsens, J. (1995). Generalization of principal component analysis, optimization problems and Neural Network. *Neural Networks*, **8** (4), pp. 549–562.
- [261] Walters-Williams, J., and Li, Y. (2011). A new approach to denoising EEG signals-merger of translation invariant wavelet and ICA. *International Journal of Biometrics and Bioinformatics*, **5** (2), pp. 130–148.
- [262] Vigario, R., Jousmaki V., Hamalainen M., Hari R., and Oja E. (1998). Independent component analysis for identification of artifacts in magnetoencephalographic recordings. *In: Jordan, M.I., Kearns, M.J., and Solla, S.A., eds. Advances in neural information processing systems*, 1998. Cambridge: MIT Press.
- [263] Inuso, G., Foresta, B., Mammone, N., and Morabito, C. F. (2007). Wavelet-ICA methodology for efficient artifact removal from Electroencephalographic recordings. *In: Proceedings of International Joint Conference on Neural Networks*, Orlando, 12-17 August 2007. Florida: IEEE.
- [264] Senthil Kumar, P., Arumuganathan, R., Sivakumar, K., and Vimal, C. (2008). Removal of ocular artifacts in the EEG through wavelet transform without using an EOG reference channel. *International Journal of Open Problems in Computer Science*, **1** (3), pp. 188–200.
- [265] Pal, P. R., and Panda, R. (2010). Classification of EEG signals for epileptic seizure evaluation. *In: proceeding of the IEEE students technology symposium*, India, Kharagpur 3-4 April 2010. Kharagpur: IEEE, pp. 72–73.
- [266] Yucel, Z., and Ozguler, A. B. (2008). Detection of epileptic indicators on clinical subbands of EEG. *In: Proceedings of the 16th European Signal Processing Conference*, Lausanne, Switzerland, 25-29 August 2008. Lausanne: IEEE, pp. 1–5.

- [267] Choe, S. H., Chung, Y. G., and Kim, S. P. (2010). Statistical spectra feature extraction for classification of epileptic EEG signals. *In: Proceeding of the International Conference on Machine Learning and Cybernetics*, Qingdao, China 11-14 July 2010. Qingdao, pp. 3180–3185.
- [268] Ravish, D. K., Shenbaga Devi, S., Krishnamoorthy, S. G., and Karthikeyan, M. R. (2013). Detection of epileptic seizure in EEG recordings by spectral method and statistical analysis. *Journal of Applied Sciences*, **13** (2), pp. 207–219.
- [269] Maasoumi, E., and Racine, J. (2002). Entropy and predictability of stock market returns. *Journal of Econometrics*, **107**, pp. 291–312.
- [270] Granger, C., and Lin, J. (1994). Using the mutual information coefficient to identify lags in nonlinear models. *Journal of Time Series Analysis*, **15**, pp. 371–384.
- [271] Darbellay, G., and Wuertz, D. (2000). The entropy as a tool for analysing statistical dependences in financial time series. *Physica A*, **287**, pp. 429–439.
- [272] Menezes, R., Dionsio, A., and Hassani, H. (2012). On the globalization of stock markets: An application of vector error correction model, mutual information and singular spectrum analysis to the G7 countries. *Quarterly Review of Economics and Finance*, **52**, pp. 369–384.
- [273] Dionisio, A., Menezes, R., and Mendes, D. A. (2006). Entropy-based independence test. *Nonlinear Dynamics*, **44**, pp. 351–357.
- [274] D’Agostino, R. B. (1986). Tests for normal distribution. *In: D’Agostino, R. B., and Stephens, M. A., eds. Goodness-of-fit techniques*. New York: Macel Dekker.
- [275] Thode, H. C. (2002). Testing for Normality. New York: Marcel Dekker, Inc.
- [276] Razali, N. M., and Wah, Y. B. (2011). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics*, **2** (1), pp. 21–33.

-
- [277] D'Agostino, R. B., Belanger, A., and D'Agostino, J. R. (1990). A suggestion for using powerful and informative tests of normality. *The American Statistician*, **44**, pp. 316–321.
- [278] Armitage, P., and Colton, T. (1998). *Encyclopedia of biostatistics*. New York: Wiley.